Syllabus

1. Introduction
   - Survival data
   - Censoring mechanism
   - Application in medical field
2. Concepts and definitions
   - Survival function
   - Hazard function
3. Non-parametric approach
   - Life table
   - Kaplan-Meier survival estimate
   - Hazard function
   - Median and percentile survival time
4. Hypothesis testing
   - Overview – hypothesis, test statistics, p-values
   - Log-rank
   - Wilcoxon
   - Gehan test
5. Study design and sample size estimation
   - Overview
   - Survival sample size estimation
   - Accrual time and Study duration
6. Semiparametric model – proportional hazard model
   - Partial likelihood
   - Inference
   - Time varying covariates
   - Stratification
7. Model checking in the PH model
   - Model checking
   - Residuals
8. Parametric model
   - Parametric proportional hazard model
   - Accelerate failure model
9. Other topics
   - **Competing risk**
   - Recurrent events
   - Non-proportional hazard ratio
   - Interval censoring

Motivating Examples

- Patients undergoing hemodialysis may be at high risk for cardiovascular events
  - If interest is in the cardiac events: stroke, myocardia infraction, etc
  - Death may censor the cardiac events, possibly informatively
- All-cause mortality can have different causes
  - Cancer, COPD, cardiovascular attacks, other random reasons
  - If interested in specific causes - time to death due to cancer
  - Death from other causes may censor the death due to the specific cause
- Certain therapies may
  - Delay disease progression but not prolong life
  - Or vice versus
  - Important to discern the two

Competing Risk

- Competing risk arises
  - When only one type of events can be observed,
    - Death – can only die once
    - Time to the first event

- Relationship among the competing events
  - Independent
  - Dependent
  - Example – death due to cancer
    - Independent censoring
      - Car accident, animal attacks, suicide, …
    - Dependent Causes:
      - Death due to diabetes, infection, …

Analysis Strategies With Competing Events

- Often composite endpoints are used in clinical trials
  - Combine different types of events
  - Examples
    - Progression – free survival
      - Time from treatment to disease progression or all-cause death
    - Relapse-free survival
      - Time from treatment o disease relapse or death
    - MACE – major adverse cardiac events
      - Stroke
      - MI
      - Hospitalization due to certain cardiovascular events
      - Death
- Not sufficient when the interest is a specific event

Survival Analysis with Competing Risk

- Topics
  - Cause-specific hazard
  - Cumulative incidence function (ICF)
    - Sub-distribution

Notations

- Let T_k be the k^{th} type of event, $k = 1, \dots, K$
- The observed survival time (T, Δ),
  - $T = \min(T_1, \dots, T_K, C)$
  - $\Delta = 0, 1, \dots, K$
    - $\Delta = 0$ represent independent censoring

Cause-specific Hazard

- Recall, definition of the hazard function

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T \geq t)}{\Delta t}$$

- Similarly, the cause-specific hazard is

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t, \Delta = k | T \geq t)}{\Delta t}$$

For $k = 1, 2, \ldots, K$ and no overlapping among the different types of events

$$h(t) = \sum_{k=1}^{K} \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t, \Delta = k | T \geq t)}{\Delta t} = \sum_{k=1}^{K} h_k(t)$$

Cause-specific Hazard

- Cause-specific cumulative hazard

$$H_k(t) = \int_0^t h_k(t)dt$$

$$H(t) = \sum_{k=1}^{K} H_k(t)$$

- Recall, $S(t) = e^{-H(t)}$
- However, cause-specific survival

$$S_k(t) \neq e^{-H_k(t)}$$

Analysis Strategies for Cause-Specific Events

- When competing events are independent
  - Treating competing events as censoring,
    - Example – for death due to cancer
    - Other causes of death will be treated as censor
  - Apply what we have learned so far
    - Non-parametric
      - Kaplan-Meier
      - Log-rank for cause-specific hazard function
    - Semi-parametric to incorporate regression
      - Cause-specific hazard regression
    - Parametric

The Consequence of Dependency

- Is it possible to test independency?
  - Tsiatis (1975) argued untestable for insufficient information

- Using an example – Stroke and all-cause mortality
  - Independent censoring means the same risk rate among
    - Those whose events are observed
    - Those who are censored
  - Dependent censoring
    - All-cause mortality can be dependent censoring of stroke
      - If not die, higher risk develop stroke – a conversation with a cardiologist
      - If treated as independent censor – may underestimate the risk

Cumulative Incidence Function (CIF)

- Cumulative incidence function (CIF) for events with only one type

$$F(t) = P(T \leq t) = \int_0^t f(t)dt$$

$$\hat{F}(t) = 1 - \hat{S}(t)$$

$\hat{S}(t)$ is a K-M estimator

- Recall, $h(t) = \frac{f(t)}{S(t)}$, we have

$$F(t) = \int_0^t S(t)dH(t)$$

Cause-specific Cumulative Incidence Function

- Cause-specific CIF is defined as

$$F_k(t) = P(T \leq t, \Delta = k)$$

$$= \int_0^t S(t) dH_k(t)$$

$$= \int_0^t S(t) h_k(t) d\,\mathrm{t}$$

$$F_k(t) = \int_0^t e^{-H(t)} dH_k(t) = \int_0^t e^{-\sum_{j=1 \text{ to } K} H_j(t)} dH_k(t)$$

$$F(t) = F_1(t) + F_2(t) + \cdots + F_K(t)$$

Sub-distribution

- Sub-distribution hazard functions

$$h_k^s(\mathrm{t}) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \Delta = k | T \geq t \cup (T < t \cap \Delta \neq k))}{\Delta t}$$

- As a function of CIF

$$h_k^s(\mathrm{t}) = \frac{d}{dt} \log(1 - F_k(t))$$

- Note

$$h_k(t) \neq h_k^s(\mathrm{t})$$

Sub-distribution

- Sub-distribution function
  - It does not correspond to a true probability distribution
- Sub-distribution hazard function considers the rate of events in those subjects
  - who are either currently event-free or
  - who have previously experienced a competing events

- The sub-distribution function is for the improper random variable

$$T^* = I(\Delta = k) \times T + \{1 - I(\Delta = k)\} \times \infty$$

  - $\{1 - I(\Delta = k)\} \times \infty$ means subjects who experienced competing events are immortal

Estimate

- Using the Nelson-Aalen estimator to estimate $H_k(t)$
  - $\widehat{H}_k(t) = \sum_{j:t_j \le t} \frac{d_j}{n_j}$,

    where d_j - number of death occurred at t_j

    n_j - number of subjects survived at t_j^-
  - Treating all other types of events as censor

Estimate

- The cause-specific CIF can be estimated by
  - $\hat{F}_k(t) = \int_0^t e^{-\hat{H}(t)} d\hat{H}_k(t)$

  where $\hat{H}(t) = \sum_{k=1}^{K} \hat{H}_k(t)$

- $\hat{F}_k(t) = \sum_{t_j \le t} \hat{S}(t^-)\hat{h}_k(t_j)$

Example – Cause-specific CIF

- Suppose we have
  - Two groups, A and B
  - Two types of events $j = 1,2$
- The cause-specific hazard are shown in the table

$$h_{A1}(t) = h_{B1}(t) = 1$$

	Group A	Group B	Hazard Ratio A/B
Event Type 1	$h_{A1}(t) = 1$	$h_{B1}(t) = 1$	1
Event Type 2	$h_{A2}(t) = 1$	$h_{B2}(t) = 3$	1/3

Example – Cause-specific CIF

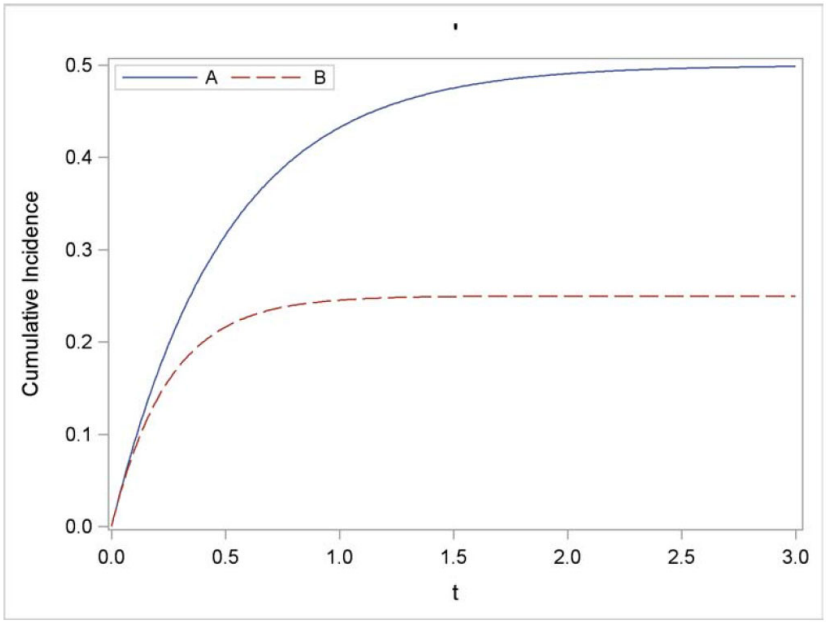
- The corresponding CIFs

$$F_{A1}(t) = \int_0^t S_A(u)h_{A1}(u)du = \int_0^t e^{-2u}du = \frac{1}{2}(1 - e^{-2t})$$

$$F_{B1}(t) = \int_0^t S_B(u)h_{B1}(u)du = \int_0^t e^{-4u}du = \frac{1}{4}(1 - e^{-4t})$$

- Notice that
  - $F_{A1}(t) \neq F_{B1}(t)$
- For Type 1 events
  - There is a group difference in CIF between groups A and B,
  - When there is no difference in cause-specific hazard functions

Figure Cumulative Incidence Functions from Groups A and B for Event 1



Fine and Gray's Model

- Introduce covariates in the context of competing risks
- Based on CIF for each type of events
- Model the sub-hazard function

$$h_k^S(t, Z) = h_{k0}^S(t) e^{\beta' Z}$$

Fine and Gray's Model

- The partial likelihood for the k^{th} type of events

$$L_k(\beta) = \prod_{j=1}^{J} \frac{e^{\beta' z_{(j)}(t_{(j)})}}{\sum_{l \in R^s(t_{(j)})} w_{lk}(t_i) e^{\beta' z_l(t_{(j)})}}$$

$$= \prod_{i=1}^{n} \left\{ \frac{e^{\beta' z_i(t_i)}}{\sum_{l \in R^s(t_i)} w_{lk}(t_i) e^{\beta' z_l(t_i)}} \right\}^{\Delta_i}$$

- Notice two differences from the regular partial likelihood
  - The definition of risk set
  - The weights included in the risk set

Fine and Gray's Model

- The risk set $R^s(t_i)$
  - All subjects survived at t_i^-
  - All subjects who had experienced a competing event before t_i

- Choice of weights w_{lk} for $l \in R^s(t_i)$
  - Subjects have not experienced a competing event before t_i
    - $w_{lk}(t) = 1$

  - Subjects experienced a competing event before t_i
    - $w_{lk}(t) = \frac{\hat{G}(t)}{\hat{G}(\min(T_l \wedge t))} < 1,$
      where $\hat{G}(t)$ is the Kaplan-Meier estimate of the survival function of the random censoring variable C
    - $w_{lk}(t)$ decreasing with time

Fine and Gray's Model

- The risk set $R^s(t_i)$ includes patients
  - Who are at risk for the event of interest
  - Who experience a competing event before t_i and are therefore *immortal*.
- The model covariates, do not directly link to the rate of the underlying event
- The interpretation of the coefficient of the model
  - Include information of the covariates on all competing events

Example – Stem Cell Transplant

- 4 : : #sdwlhqw#z kr#uhfhlyhg#d#vwhp #fho#wudqvsodqw#ru#dfxwh# dnxnhp ld#
- Wkh#hyhqw#ri#qwhuhvw#q#uhodsvh/#exw#rwkhu#frp shwlqj#fdxvhv# +h1j1wudqvsodqw0uhodwhg#ghdwk,#qhhg#wr#eh#wdnhq#qwr#dffrxqw#
- Fryduldwhv#vxfk#dv#Vh{/#G lvhdvh#o|p skreodvwlf#ru# p |horeodvwlf#dnxnhp ld/#deeuhyldwhg#dv#DOO#dqg#DP O/# uhvshfwlyho|,#Skdvh#dw#wudqvsodqw#Uhodsvh/#FU4/#FU5/#FU6,# Vrxufh#ri#vwhp #fhov#erqh#p duurz #dqg#shulskhudo#eorrg/# frghg#dv#EP .SE/#ru#shulskhudo#eorrg/#frghg#dv#SE,#dqg#Djh1#

Example – Stem Cell Transplant

Variable	Description	Statistical summary
Sex	Sex	M=Male (100) F=Female (77)
D	Disease	ALL (73) AML (104)
Phase	Phase	CR1 (47) CR2 (45) CR3 (12) Relapse (73)
Source	Type of transplant	BM+PB (21) PB (156)
Age	Age of patient (years)	4–62 30.47 (13.04)
Ftime	Failure time (months)	0.13–131.77 20.28 (30.78)
Status	Status indicator	0=censored (46) 1=relapse (56) 2=competing event (75)

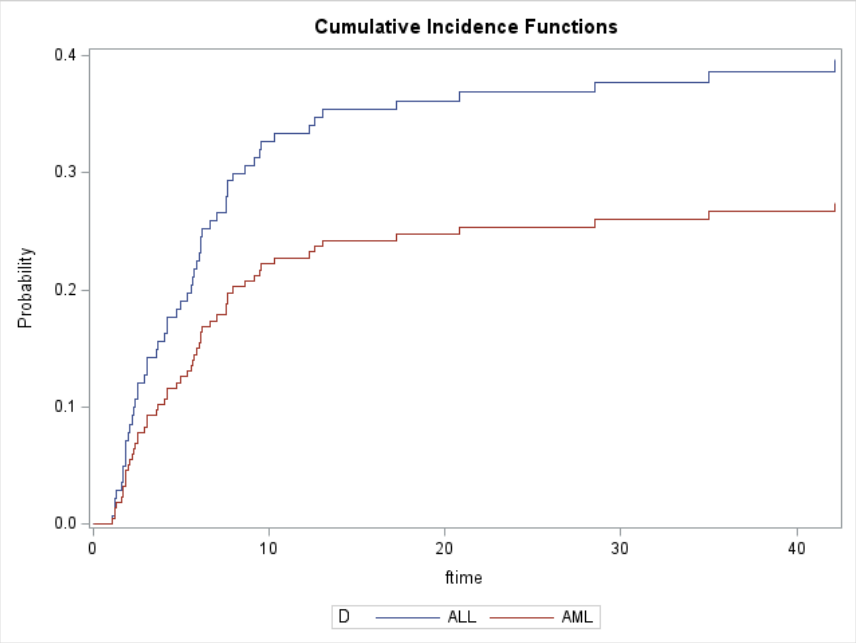
Relapse is the event of interest

D				
D	Frequency	Percent	Cumulative Frequency	Cumulative Percent
ALL	73	41.24	73	41.24
AML	104	58.76	177	100.00

Summary of Failure Outcomes			
Total	Event of Interest	Competing Event	Censored
177	56	75	46

Example – Stem Cell Transplant

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
D	AML	1	-0.45326	0.26571	2.9099	0.0880	0.636	D AML




```
data Risk;
   D="ALL"; output;
   D="AML"; output;
   format D $3.;
   run;
ods graphics on;
proc phreg data=example
plots(overlay=stratum)=c
if;
   class D
(order=internal
ref=first);
   model
ftime*Status(0)=D /
eventcode=1;
   Hazardratio
'Pairwise' D /
diff=pairwise;
   baseline
covariates=Risk out=out1
cif=_all_ / seed=99333;
run;
```

Relapse is the event of interest

Example – Stem Cell Transplant

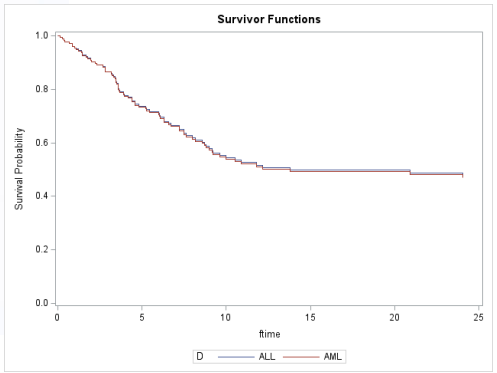
Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
177	75	102	57.63

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
D	1	0.0048	0.9450

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
D	AML	1	0.01647	0.23888	0.0048	0.9450	1.017	D AML

```
***Analyzing the competing events;
***Cause-specific for the competing events;

proc phreg data=example
plots(overlay=stratum)=survival;
   class D (order=internal ref=first);
   model ftime*Status(0,1)=D;
   Hazardratio 'Pairwise' D / diff=pairwise;
   baseline covariates=Risk out=out1 cif=_all_ /
seed=99333;
run;
```



Example – Stem Cell Transplant

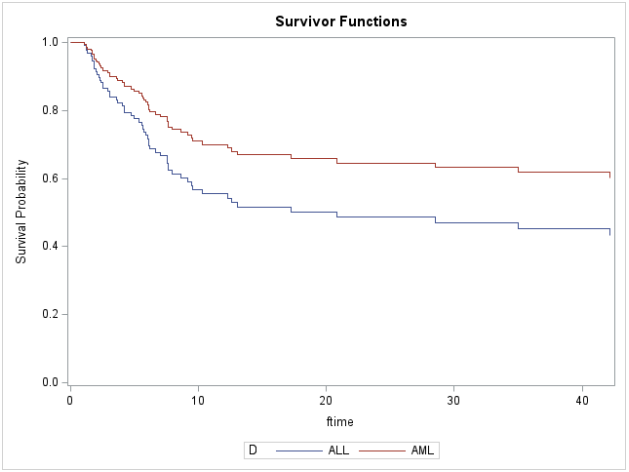
Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
177	56	121	68.36

```
***Analyzing the events of interest;
***Cause-specific for relapse;

proc phreg data=example
plots(overlay=stratum)=survival;
   class D (order=internal ref=first);
   model ftime*Status(0,2)=D;
   Hazardratio 'Pairwise' D / diff=pairwise;
   baseline covariates=Risk out=out1 cif=_all_ /
seed=99333;
```

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
D	1	3.4943	0.0616

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
D	AML	1	-0.49980	0.26737	3.4943	0.0616	0.607	D AML



Example – Stem Cell Transplant

Fine-Gray Method

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Wald	4.8468	3	0.1834

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
D	1	0.9749	0.3234
Age	1	2.4448	0.1179
Sex	1	0.0059	0.9389

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
D	AML	1	-0.27549	0.27901	0.9749	0.3234	0.759	D AML
Age		1	-0.01836	0.01174	2.4448	0.1179	0.982	Age
Sex	F	1	-0.02102	0.27443	0.0059	0.9389	0.979	Sex F

Cause-specific

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.7899	3	0.1878
Score	4.8030	3	0.1868
Wald	4.7105	3	0.1943

Type 3 Tests			
Effect	DF	Wald Chi-Square	Pr > ChiSq
D	1	2.0187	0.1554
Sex	1	0.5097	0.4753
Age	1	1.0069	0.3156

Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
D	AML	1	-0.40720	0.28660	2.0187	0.1554	0.666	D AML
Sex	F	1	0.19777	0.27701	0.5097	0.4753	1.219	Sex F
Age		1	-0.01175	0.01171	1.0069	0.3156	0.988	Age

Non-parametric Test - The Gray's Test (1988)

- Let T_{ki} be the k^{th} type of event in the i^{th} group,
  - $k = 1, \dots, K$ and $i = 1, 2, \dots, \mathrm{I}$
- H_0: The cause-specific CIF are identical across treatment groups

  $F_{k1}(t) = F_{k2}(t) =, \dots, = F_{kI}(t)$
- H_A: The cause-specific CIF are not all the same

- Note, the test is for the sub-distribution
  - CIF
  - hazard

The Gray's Test

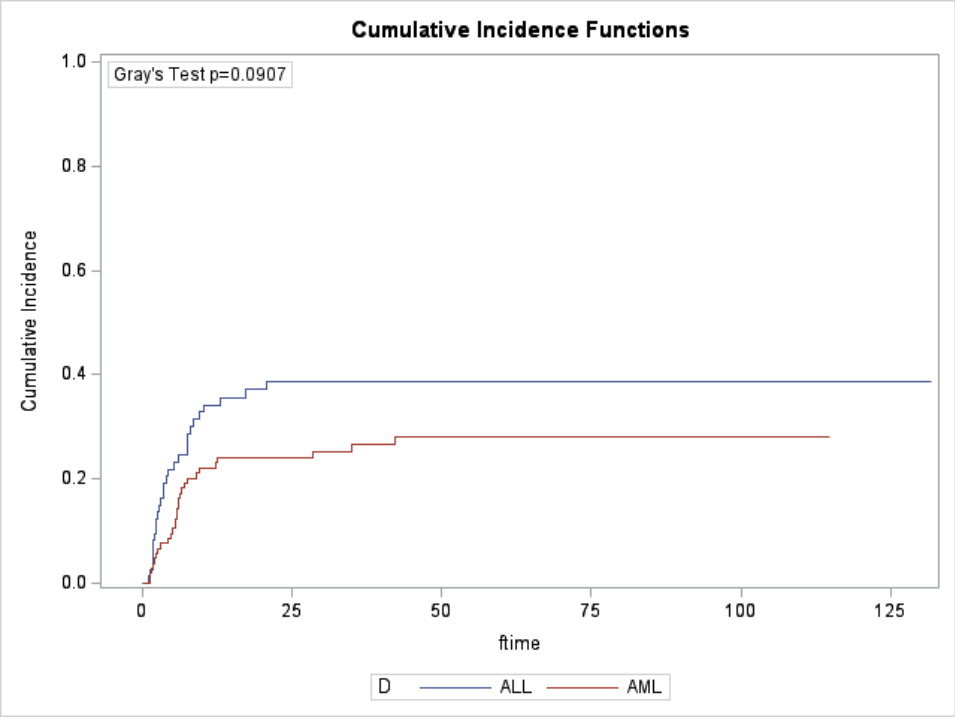
- Let T_{ki} be the k^{th} type of event in the i^{th} group,
  - $K = 2$ and $I = 2$

- $\int_0^{\tau} W(t) \left\{ \frac{d\hat{F}_{11}(t)}{1-\hat{F}_{11}(t^-)} - \frac{d\hat{F}_{12}(t)}{1-\hat{F}_{12}(t^-)} \right\}$

- Basically, the test statistics compares weighted averages of the "sub-distribution hazards"

$$\frac{f_{1i}}{1 - F_{1i}}$$

Example – Stem Cell Transplant




```
***Gray Test;
ods graphics on;
proc lifetest data=example
plots=cif(test);
   time ftime*Status(0)/eventcode=1;
   strata D / order=internal;
run;
```

Cumulative Incidence Function Estimates

Stratum 2: D = AML

ftime	Cumulative Incidence	Standard Error	95% Confidence Interval	
0	0	0	.	.
1.2	0.00962	0.00962	0.000839	0.0476
1.3	0.0192	0.0135	0.00369	0.0616
1.6	0.0288	0.0165	0.00772	0.0754
1.87	0.0385	0.0190	0.0125	0.0887
2.03	0.0481	0.0211	0.0178	0.1016
2.3	0.0577	0.0230	0.0235	0.1142
2.53	0.0673	0.0247	0.0295	0.1265
3.03	0.0769	0.0263	0.0358	0.1386
4.2	0.0865	0.0277	0.0423	0.1506

Homework 10

1. Write down the definition of the overall hazards, cause specific hazards, and sub-distribution hazards. Describe the relationship and differences.
2. Use the melanoma data to analyze the effect of sex on the cause specific death from melanoma. Note, status=1,2,3:

   1=died of melanoma

   2=alive

   3=died from other reasons
3. Use the melanoma data again to test the sex effect using sub-distribution hazards.