

Levetidsanalyse (survival analysis)
Rosner 14.8-14.11

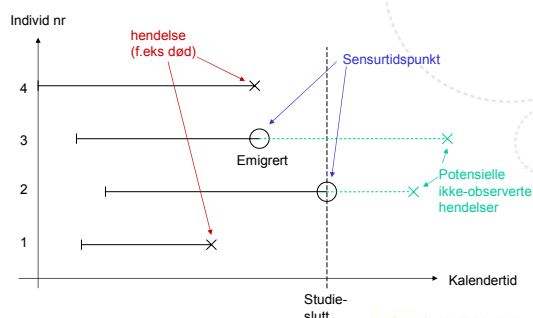
av Stian Lydersen

Forlesning 16 april 2008

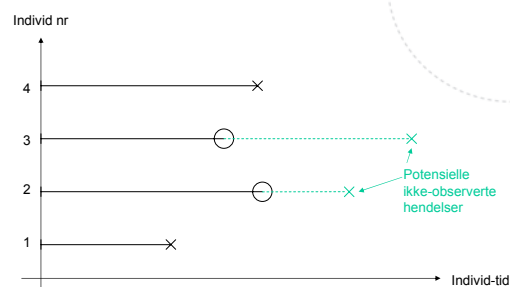
Levetid (varighet av en tilstand)

- Eksempler:
 - Tid til personen dør (målt fra fødsel, fra diagnose, fra behandling)
 - Tid til en diagnose
 - Tid fra start av behandling til erklært frisk
- Levetider er
 - Alltid større enn 0
 - Ofte sensurert
- $S(t)$ = Sannsynligheten for å overleve t

Sensurerte data



Sensurerte data



Levetidsanalyse - mest brukte metoder

- Aktuarberegning (grov beregning innenfor tidsintervall)
- Kaplan-Meier plott (empirisk overlevelsessannsynlighet)
- Log-rank test: Ikke-parametrisk test for forskjell mellom grupper
- Regresjonsanalyse:
 - Semi-parametrisk (Cox regresjon = Proporsjonal hazard regresjon)
 - Parametrisk (Antar bestemt form på fordelingen, f.eks Weibull-fordeling)

ISI-søk 10 April 2008

COX DR
REGRESSION MODELS AND LIFE-TABLES
JOURNAL OF THE ROYAL STATISTICAL SOCIETY
SERIES B-STATISTICAL METHODOLOGY 34 (2):
187& 1972
Times Cited: 22873

Title: NONPARAMETRIC-ESTIMATION FROM
INCOMPLETE OBSERVATIONS
Author(s): **KAPLAN EL, MEIER P**
Source: JOURNAL OF THE AMERICAN STATISTICAL
ASSOCIATION 53 (282): 457-481 1958

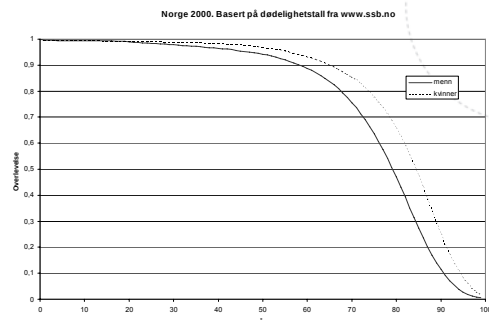
Times Cited: 32059

7

Metodene for levetidsanalyse forutsetter ikke-informativ sensurering

- For hvert individ finnes en potensiell levetid T og et sensurtidspunkt C .
- Observerer kun $\min(T, C)$ og D hvor
 - $D=0$ hvis sensurert
 - $D=1$ hvis hendelse (f.eks død)
- Metodene for levetidsanalyse forutsetter at T, D er uavhengige
- Umulig å sjekke uavhengighet fra data

8



9

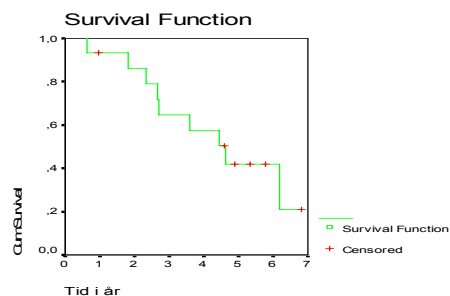
Kaplan-Meier (produkt-grense) estimatoren for overlevelsessannsynligheten

Data om overlevelse for 15 pasienter med melanom (fra Aalen, 1998)

Antall i live og under observasjon	Status (0=sensurert, 1=død)	Observasjons-tid i år	Overlevelses-sannsynlighet (Kaplan-Meier)
15	1	0,64	0,933
14	0	0,97	
13	1	1,81	0,862
12	1	2,35	0,79
11	1	2,65	0,718
10	1	2,69	0,646
9	1	3,59	0,574
8	1	4,44	0,503
7	0	4,6	
6	1	4,63	0,419
5	0	4,9	
4	0	5,32	
3	0	5,76	
2	1	6,18	0,209
1	0	6,83	

10

Kaplan-Meier plott (empirisk overlevelsessannsynlighet)



11

Sammenlikning mellom to grupper

H_0 : Ingen gruppeeffekt
 $S_1(t) = S_2(t)$ for alle t

mot

H_1 : Levetidene tenderer til å være lengre (evt kortere) i gruppe 1
 $S_1(t) \geq S_2(t)$ for alle t med $>$ for noen t (evt $\leq$ og $<$)

Ikke-parametrisk test for komplette eller sensurerte data:
 Lograng-test.

Tilnærmet samme resultat:
 Test om $\beta=0$ i Cox modell med x som indikator på gruppe.
 (Eksakt samme resultat hvis ingen sammenfallende observasjoner OG bruk av LR p-verdi i Cox modell.)

12

Kleinbaum & Klein (2005) s 17
 Remisjonstider for 42 leukemi-pasienter (Freireich & al, Blood, 1963)

uker	hend.	grp	uker	hend.	grp	uker	hend.	grp
1	1	2	8	1	2	16	1	1
1	1	2	8	1	2	17	0	1
2	1	2	8	1	2	17	1	2
2	1	2	8	1	2	19	0	1
3	1	2	9	0	1	20	0	1
4	1	2	10	1	1	22	1	1
4	1	2	10	0	1	22	1	2
5	1	2	11	0	1	23	1	1
5	1	2	11	1	2	23	1	2
6	1	1	11	1	2	25	0	1
6	1	1	12	1	2	32	0	1
6	1	1	12	1	2	32	0	1
6	0	1	13	1	1	34	0	1
7	1	1	15	1	2	35	0	1

Hendelse
 0 = sensur
 1 = remisjon
 Grp
 1= Behandling
 2=placebo

13

Lograng-test:
Eksempel fra
Kleinbaum & Klein
(2005)

Tid (uke)	hendelser		rest	
	gr1 h1	gr2 h2	gr1 r1	gr2 r2
1	0	2	21	21
2	0	2	21	19
3	0	1	21	17
4	0	2	21	16
5	0	2	21	14
6	3	0	21	12
7	1	0	17*	12
8	0	4	16	12
10	1	0	15	8
11	0	2	13	8
12	0	2	12	6
13	1	0	12	4
15	0	1	11	4
16	1	0	11	3
17	0	1	10	3
22	1	1	7	2
23	1	1	6	1
Sum	9	21		

* Sensurerte tider
kommer ikke frem,
kursiv antyder
uker med sensur.

NTNU
Det skapende universitet

www.ntnu.no

14

Log-rang testen

Likner på en kjiqkvadrat "goodness of fit" test
Del opp i korte tidsintervall
(bare ett tidspunkt med hendelse i hvert intervall)

Forventet antall hendelser i intervall nr j:
Expected = andel i risikomengden × antall hendelser totalt

$$e_{1j} = \left(\frac{r_{1j}}{r_{1j} + r_{2j}} \right) \times (h_{1j} + h_{2j}),$$

$$e_{2j} = \left(\frac{r_{2j}}{r_{1j} + r_{2j}} \right) \times (h_{1j} + h_{2j})$$

NTNU
Det skapende universitet

www.ntnu.no

15

Tid (uke)	hendelser		rest		expected		Gr2** Obs-Exp
	gr1 h1	gr2 h2	gr1 r1	gr2 r2	gr1 e1	gr2 e2	
1	0	2	21	21	21/42 *2	21/42 *2	1.00
2	0	2	21	19	21/40 *2	19/40 *2	1.05
3	0	1	21	17	21/38 *1	17/38 *1	0.55
4	0	2	21	16	21/37 *2	16/37 *2	1.14
5	0	2	21	14	21/35 *2	14/35 *2	1.20
6	3	0	21	12	21/33 *3	12/33 *3	-1.09
7	1	0	17*	12	17/29 *2	12/29 *2	-0.41
8	0	4	16	12	16/28 *4	12/28 *4	2.29
10	1	0	15	8	15/23 *1	8/23 *1	-0.35
11	0	2	13	8	13/21 *2	8/21 *2	1.24
12	0	2	12	6	12/18 *2	6/18 *2	1.33
13	1	0	12	4	12/16 *1	4/16 *1	-0.25
15	0	1	11	4	11/15 *1	4/15 *1	0.73
16	1	0	11	3	11/14 *1	3/14 *1	-0.21
17	0	1	10	3	10/13 *1	3/13 *1	0.77
22	1	1	7	2	7/9 *2	2/9 *2	0.56
23	1	1	6	1	6/7 *2	1/7 *2	0.71
Sum	9	21			19.26	10.74	10.3

sensurerte kommer ikke frem, kursiv antyder uker med sensur
*Observert-expected for gruppe 1 = -(Observert-expected) for gruppe 2

NTNU
Det skapende universitet

www.ntnu.no

16

Log-rank test

- log-rank observator: $(\sum_{\text{failure times}} \text{obs-forv})^2 / \text{var} (\sum \text{obs-forv})$
- var $(\sum \text{obs-forv}) = \sum \frac{r1 * r2 * (h1 + h2) * (r1 + r2 - h1 - h2)}{(r1 + r2)^2 * (r1 + r2 - 1)} = 6.3$
- Når to grupper blir sammenliknet brukes kun resultater for en gruppe; (obs-forv) for gruppe 1 = - (obs-forv) for gruppe 2
- For gruppe 2: $(\sum_{\text{failure times}} \text{obs-forv})^2 = (10.3)^2 = 106.09$
- log-rank: $\chi^2 = 16.8$, df = 1 gir p-verdi < 0.0001
- Mer komplisert hvis det er mer enn 2 grupper som sammenliknes

NTNU
Det skapende universitet

www.ntnu.no

17

Overlevelsessannsynlighet (survival probability, survival function):

$$S(t) = P(T > t)$$

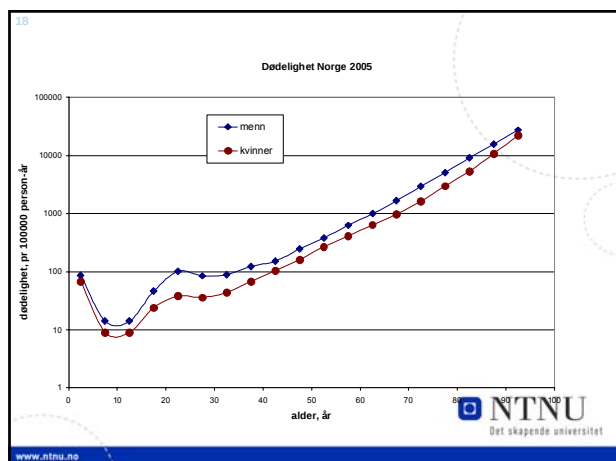
Hazard rate, dødelighet (force of mortality): $h(t)$

Sannsynligheten for at individet av "alder" t "dør" i løpet av neste tidsintervall av varighet Δt :

$$P(T \leq t + \Delta t \mid T > t) \approx h(t) \Delta t$$

NTNU
Det skapende universitet

www.ntnu.no



19

Formell definisjon av hazardraten:

$$z(t) = \lim_{\Delta t \rightarrow 0} \frac{P(T \leq t + \Delta t | T > t)}{\Delta t}$$

som vi ser blir lik

$$z(t) = \lim_{\Delta t \rightarrow 0} \frac{P((T \leq t + \Delta t) \cap (T > t))}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t)}{\Delta t} \frac{1}{P(T > t)}$$

$$= \lim_{\Delta t \rightarrow 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} \frac{1}{P(T > t)} = \frac{f(t)}{S(t)}$$

og også:

$$z(t) = -\frac{d}{dt}(\ln(S(t)))$$

NTNU
Det skapende universitet

www.ntnu.no

20

Sammenheng mellom S(t) og h(t):

$$S(t) = e^{-\int_0^t h(u) du} = e^{-H(t)}$$

hvor

$$H(t) = \int_0^t h(u) du$$

kalles kumulativ hazard.

Det er lettere å plote H(t) enn h(t).

Hazardraten h(t) = H'(t) er stigningstallet til H(t).

NTNU
Det skapende universitet

www.ntnu.no

21

Cox Proporsjonal Hazard (PH) modell

$$h(t; \underline{x}) = h_0(t) \exp(\beta \underline{x})$$

$$h(t; x_1, \dots, x_p) = h_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p)$$

$$= h_0(t) \exp(\beta_1 x_1) \dots \exp(\beta_p x_p)$$

ekvivalent:

$$S(t; \underline{x}) = S_0(t)^{\exp(\beta \underline{x})}$$

hvor $h_0(t)$, $S_0(t)$ kalles "base line hazard rate",
"base line overlevelsessannsynlighet"

$h_0(t)$ er ukjent og helt uspesifisert
 $\beta_1, \dots, \beta_p$ er ukjente parametre

NTNU
Det skapende universitet

www.ntnu.no

22

En fortolkning av β

- Hvis to individer (eller grupper) har hhv $x_1=0$ og $x_1=1$ og ellers like forklaringsvariable: Forholdet mellom hazardratene for gruppene er lik $\exp(\beta_1)$, uansett "alder" t.
- Størrelsen $\exp(\beta_1)$ kalles hazard ratio (HR).
- Relevant bare hvis forklaringsvariabelen ikke inngår i interaksjonsledd!

NTNU
Det skapende universitet

www.ntnu.no

23

Cox, D. R. (1972):

- Metode for estimering av $\beta_1, \dots, \beta_p$
- og ikke-parametrisk estimering av $S_0(t)$ (Kaplan-Meier type estimering)
- for komplette og sensurerte datasett
- Tidsavhengige $\beta_i(t)$ er mulig
- (Alternativt: Parametrisk modell, f.eks lognormal eller Weibullfordeling for S(t).)

NTNU
Det skapende universitet

www.ntnu.no

24

Partiell likelihood (Cox):

$$l_p(\underline{\beta}) = \prod_{i=1}^m \frac{h_0(t_i) \exp(\underline{x}_{(i)} \underline{\beta})}{\sum_{j \in R(t_i)} h_0(t_i) \exp(\underline{x}_{(j)} \underline{\beta})} = \prod_{i=1}^m \frac{\exp(\underline{x}_{(i)} \underline{\beta})}{\sum_{j \in R(t_i)} \exp(\underline{x}_{(j)} \underline{\beta})}$$

hvor $t_{(1)} < t_{(2)} < \dots < t_{(m)}$ er de m distinkte (usensurerte) levetidene.

$R(t_{(j)})$ er risikomengden ved $t_{(j)}$ dvs individer i live og usensurert ved $t_{(j)}$

$\underline{x}_{(j)}$ er kovariatene til individet med levetid $t_{(j)}$.

NTNU
Det skapende universitet

www.ntnu.no

25

Den partielle likelihood er ikke avhengig av $h_0(t)$!

Estimater for $\underline{\beta}$ samt estimatorenes varianser og kovarianser finnes ved å behandle $l_p(\underline{\beta})$ som en vanlig likelihood.

26

Ved sammenfallende observasjoner brukes en mer generell $l_p(\underline{\beta})$.

Alternativer:

- eksakt (meget tidkrevende)
- Breslow (1974)
- Efron (1977) (Nesten identiske resultater som eksakt).

SPSS: Bare Breslow er tilgjengelig.

Stata: Alle 3 tilgjengelig. Breslow er default.

Angi option "Efron" ved sammenfallende observasjoner

27

Sjekking av PH-antakelsen: Log - log plott

Vi har

$$S(t, \underline{x}) = S_0(t)^{\exp(\mathbf{x}'\underline{\beta})}$$

så

$$\log(-\log(S(t, \underline{x}))) = \mathbf{x}'\underline{\beta} + \log(-\log(S_0(t)))$$

Plott for forskjellige $\underline{x}$ bør være parallelle under PH antakelsen.
Vanskelig å vurdere parallellitet når det er få observasjoner.

28

Software for Cox PH regresjon

- SPSS - har endel
- MINITAB – har endel
- STATA - har det meste
- S-plus eller R- har det meste
- SAS - har det meste

29

Referanser

- Hosmer, D. W., Lemeshow, D. W., May, S.: "Applied Survival Analysis. Regression Modeling of Time to Event Data." 2nd Ed. Wiley, 2008.
 - Omfattende og velskrevet bok.
- Kleinbaum, D. G.: "Survival Analysis: A Self-Learning Text". 2nd ed. Springer, 2005.
 - Lettest innføringsbok