

Faculty of Health Sciences

Basal statistik

Overlevelsesanalyse i SPSS

Lene Theil Skovgaard

11. maj 2020



large Overlevelsesanalyse

- ▶ Levetider og censurerede observationer
- ▶ Kaplan-Meier kurver
- ▶ Log-rank test
- ▶ Cox regression
- ▶ Modelkontrol og udvidede modeller

Home pages:

<http://publicifsv.sund.ku.dk/~sr/BasicStatistics>

E-mail: ltsk@sund.ku.dk

*: Siden er lidt teknisk



Forklarende variable = Kovariater

Outcome

Outcome	Dikotom	Kategorisk	Kvantitativ	Kategoriske og kvantitative
Dikotom parret	2*2-tabeller Mc Nemar	χ^2 -test svært, mixed models		Logistisk regression Mixed models
Kategorisk	Kontingenstabeller/ χ^2 -test		Generaliseret logistisk regression	
Ordinale	svært, f.eks. proportional odds modeller			
Kvantitativ parret	Mann-Whitney Wilcoxon signed rank	Kruskal-Wallis Friedman		Robust multipel regression
Normalfordelte residualer	T-test uparret/parret	Variansanalyse ensidet/tosidet		Kovariansanalyse Multipel regression
Censureret	Log-rank test		Cox regression	
Korrelerede kvantitative Nf. residualer	Varianskomponent-modeller Mixed models		Modeller for gentagne målinger	



Eksempel: Lungecancer blandt krigsveteraner

137 krigsveteraner med diagnosen inoperabel lungecancer, af **4 forskellige typer** (1:squamous, 2:smallcell, 3:adeno, 4:other) følges frem til død (vi glemmer lige et øjeblik, at 9 af dem faktisk endnu er i live ved tidspunktet for analysen)

De er randomiseret til enten

- ▶ standard behandling ($\text{treat}=0$, $n=69$)
- ▶ ny test behandling ($\text{treat}=1$, $n=68$)

Vi vil gerne sammenligne disse to behandlinger ...og så har vi nogle kovariater (udover behandling og type), som vi vender tilbage til



Valg af tidsskala

Der skal vælges en *tidsskala* for levetiden, dvs. et nulpunkt.

Typiske muligheder er:

- ▶ alder
- ▶ kalendertid
- ▶ tid siden diagnose
- ▶ **tid siden randomisering** (vælges her)

Man kan f.eks.

- ▶ få diagnosen som 40-årig, indgå i undersøgelsen fra dette tidspunkt og dø som 72-årig....
- ▶ få diagnosen som 50-årig, randomiseres til undersøgelsen som 55-årig og dø som 65-årig....



Valg af tidsskala, fortsat

Ofte er der flere muligheder, og **valget er ikke altid helt oplagt**:

Der skal være en grund til at tro, at den valgte tidsskala afspejler risikoen ens for alle individer (på nær diverse kovariater), altså at der som funktion af tiden ses samme *form* på risikoen.

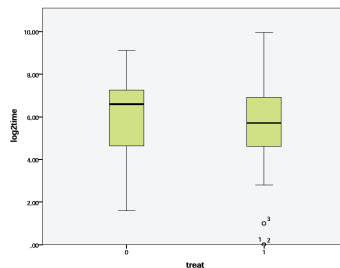
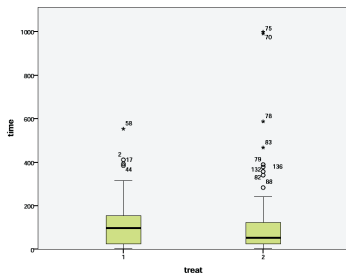
Eksempelvis kunne man vælge

- ▶ For kroniske sygdomme:
Alder: Fordi behandlingen formodentlig ikke ændrer radikalt ved prognosen
- ▶ For stærkt behandlingskrævende sygdomme:
Tid siden behandling: iden man kan formode, at en vellykket behandling på afgørende vis ændrer prognosen



Sammenligning af behandlinger

Hvis alle var fulgt indtil død, kunne vi lave T-test el.lign.



T-test på log-skala: $P=0.50$,

Mann-Whitney: $P=0.35$

men men men.....

Hvad er nyt?

Outcome: Tid fra et veldefineret nulpunkt og indtil en bestemt begivenhed (sædvanligvis død) indtræffer.

Her: Levetid i dage siden randomisering, angivet som y_i .

Det er her, der sker noget nyt, fordi det ikke er alle, der er døde endnu

Vi kalder det for censurerede observationer

Her er der kun 9 af sådanne, men ofte er der mange flere

Kovariater: Behandling, cancertype, alder, etc. etc.

Ikke noget *essentielt* nyt her....



Notation for outcome

t_i : Det tidspunkt (regnet fra nulpunktet),
der er tilknyttet det i 'te individ,
og som angiver *enten* død *eller* censureringstidspunkt.

d_i : Status ved tid t_i :

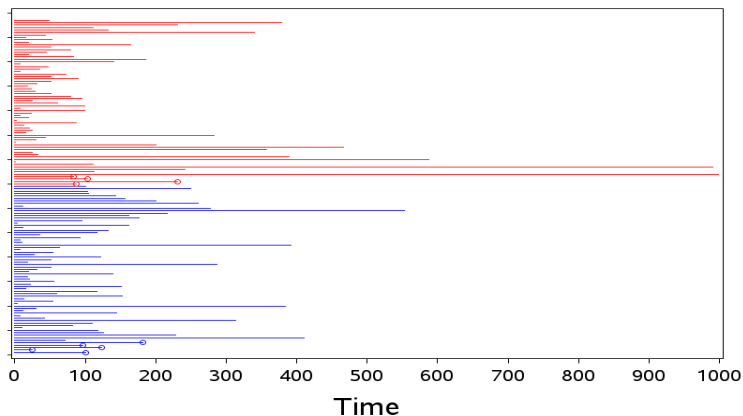
- ▶ 0: censureret, dvs. stadig i live
- ▶ 1: død

y_i : Tid til død for i 'te patient er så ikke observeret for alle:

- ▶ Hvis en patient dør til tid t_i , så er $y_i = t_i$ og $d_i = 1$
- ▶ Hvis man mister kontakten til en patient til tid t_i ,
så er $y_i > t_i$ og $d_i = 0$
- ▶ Hvis analysen laves til tid t_i , *inden* patient i dør,
så er $y_i > t_i$ og $d_i = 0$



Illustration af data



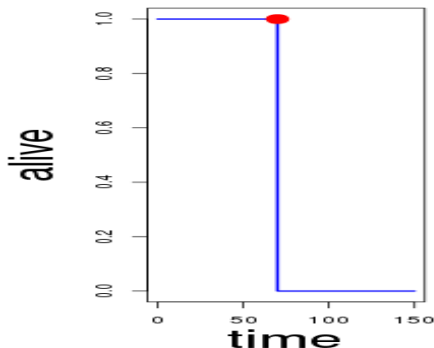
treat=0, treat=1

Cirklene angiver censurerede observationer



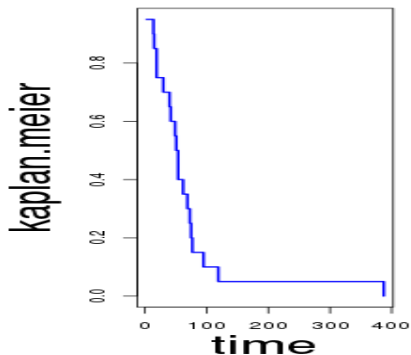
Hvad ved vi?

Vi vil opfatte tiden som *dynamisk*, i den forstand, at vi for hvert tidspunkt t ser på, hvilken information vi har om den enkelte, nemlig status for overlevelse: $S_i(t) = I(y_i > t)$



Kaplan-Meier: Overlevelseskurve

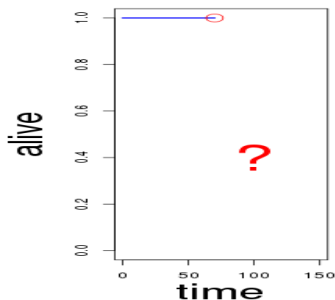
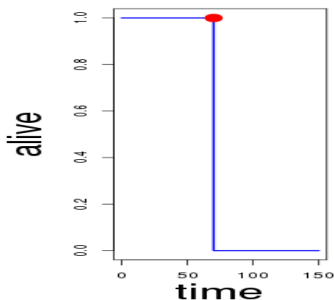
Hvis alle er observeret til død, er overlevelseskurven $S(t)$ blot gennemsnittet af disse, altså til hvert tidspunkt blot fraktionen, der stadig er i live. Den kaldes Kaplan-Meier kurven



Men når vi har censureringer?

så ved vi ikke, hvad der sker med funktionen $S_i(t) = I(y_i > t)$ efter det tidspunkt t_i , hvor vi sidst ser patienten i live.

Vi ved så kun, at deres levetid er større end t_i ,
men ikke *hvor meget* større.



Vigtigt - Uafhængige censureringer?

En **meget vigtig** antagelse om censureringer:

De skal være **uafhængige** af prognosen, dvs.

Det faktum, at en patient censureres ved tid t_i må på *ingen måde* være relateret til den fremtidige risiko for at dø

Censureringer, der muligvis *ikke* opfylder dette kriterium:

- ▶ Patienter, der overflyttes til et andet hospital:
De har formentlig en dårligere prognose
- ▶ Patienter, der flytter udenlands pga nyt job:
De har formentlig en bedre prognose

Og hvad, hvis

- ▶ patienten bliver kørt over?



Kaplan-Meier kurver, I

Når der er censureringer, kan vi ikke tage gennemsnittet af de individuelle S_i 'er, fordi vi ikke kender dem alle.

Men hvis censureringerne er uafhængige, kan vi estimere den del, vi ikke kender til blot at være "*ligesom dem, vi ser*", altså antage, at disse patienter dør i samme takt som de øvrige

Det er netop det, der ligger i uafhængig censurering.

Så hvis det, at man bliver kørt over, tyder på, at man har det dårligt (har fået recidiv), så har man nok på dette tidspunkt en værre prognose end de, der *ikke* blev kørt over....?



Håndtering af censureringer

Hvorfor kan vi ikke bare...

- ▶ Smide de censurerede individer helt væk?

Vi kender jo ikke deres levetid, så der er vel tale om *missing values*?

Nej, for vi ved *noget* om deres levetid, nemlig at den er mindst t_i , og hvis vi smider dem ud, laver vi derfor *bias* i overlevelseskurven

- ▶ Lade som om alle t_i 'er er dødsfald?

Fordi så ser det jo ud som om de dør alt for hurtigt, så her bliver også *bias*

Vi er nødt til at håndtere censureringer på en *dynamisk* måde, dvs. ved at se på situationen til et givet tidspunkt og et kort tidsrum herefter.



Kaplan-Meier kurver, II

når der er censureringer:

- ▶ Vi deler tidsaksen op ved hvert tidspunkt $t_1 < t_2 < t_3 \dots$ (her sker der noget, enten et dødsfald, eller en censurering)
- ▶ For hvert interval, f.eks. (t_3, t_4) ser vi på, hvor mange, der var *i risiko* i starten (n_3) og hvor stor en del af disse, der overlevede tidsintervallet, nemlig brøken

$$R_4 = \begin{cases} 1 & \text{hvis } t_4 \text{ er en censurering} \\ \frac{n_3-1}{n_3} & \text{hvis } t_4 \text{ er et dødsfald} \end{cases}$$

- ▶ Kaplan-Meier kurven fås nu ved successiv sammengangning af disse brøker:

$$1 \times R_1 \times R_2 \times R_3 \times R_4 \times \dots$$



Praktisk håndtering af censurering

Til programmering er outcome 2-delt:

- t_i Observationstiden for i 'te patient (time)
som kan være *enten* død (status=1), *eller* en
censurering (status=0).
- d_i Indikator for død (status), dvs. 1, hvis patienten
faktisk er død, og 0, hvis vedkommende er censureret

Begge disse variable skal angives som (todelt) outcome:

- ▶ time er det egentlige outcome,
men
- ▶ status er vigtig pga censureringerne



Kaplan-Meier kurver i SPSS

tegnes med `Analyse/Survival/Kaplan-Meier`, hvor vi sætter

- ▶ `time` over i feltet `Time`
- ▶ `status` i feltet `Status`, tryk herefter på `Define Event` og skriv 1 i `Single Value`, herefter `Continue`
- ▶ `treat` i `Factor`

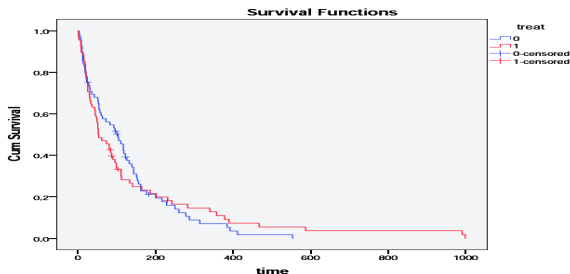
Gå derefter ind i `Options`, og sørg for, at der under `Plots` er afkrydset i `Survival` (men fjern til gengæld fluebenene i `Survival table(s)` og `Mean and median survival`).

Hvis man vil have konfidensgrænser på, er man nødt til at gøre et stort stykke arbejde selv, men det er muligt ved hjælp af `Save`-knappen at gemme standard error på disse overlevelseskurver. Figuren ses s. 20



Kaplan-Meier kurver, $S(t)$

fra Analyse/Survival/Kaplan-Meier, se forrige side:

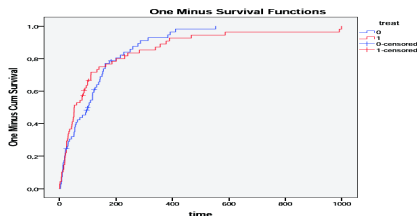


Til ethvert tidspunkt kan vi aflæse den forventede procentdel af patienter, der stadig er i live - for hver behandlingsgruppe

Kumulativ incidens

Til et hvert tidspunkt angiver dette sandsynligheden for at have fået et event, dvs. (i dette tilfælde) at være død. Denne er blot

$$CI(t) = 1 - S(t)$$



Figuren dannes via Analyse/Survival/Kaplan-Meier (se mere s. 86). **Pas på i tilfælde af *competing risks*** (se s. 81-82)



Er der forskel på de to behandlinger?

De to overlevelseskurver (s. 20) krydser lidt frem og tilbage over hinanden. De eksperimentelt behandlede dør hurtigst til at starte med, men efter ca. 100 dage ser det ud til at vende (bedømt på hældningen af kurverne).

Hvordan kan vi sammenligne?:

- ▶ **log-rank test:**
giver en P-værdi for test af identitet
- ▶ **Cox proportional hazard model:**
giver estimat for forskel (ratio) mellem dødsrater
– under *forudsætning* af proportionale rater
(kommer om lidt...)



Log-rank testet for H_0 : ens dødelighed

Et **non-parametrisk** test til sammenligning af overlevelseskurver **uden antagelser om formen af disse**

- ▶ Til hvert tidspunkt ser man på forskellen mellem observeret og forventet antal døde (under H_0) i de to grupper.
- ▶ Forskellen mellem disse, passende normeret, giver et χ^2 -test med $G - 1$ frihedsgrader (G er antallet af grupper, her $G = 2$)
- ▶ Det er ikke så stærkt, hvis forskellen på kurverne ændrer sig markant over tid.

Her giver testet $P = 0.93$ (se s. 25), altså **ingen påviselig forskel**



Log-rank testet i praksis

Testet fås som et såkaldt *score*-test (*teknisk*: det har at gøre med tangenthældningen på likelihoodfunktionen).

Vi benytter Analyse/Survival/Cox Regression, hvor vi sætter

- ▶ time over i feltet Time
- ▶ status i feltet Status, tryk herefter på Define Event og skriv 1 i Single Value, herefter Continue
- ▶ treat over i Covariates
(evt. står der Block 1 of 1, fordi man kan lave flere modeller i samme kørsel, se s. 94 og evt. s. 70)

Log-rank testet aflæses nu under Overall(score)



Output fra Log-rank test

Case Processing Summary

		N	Percent
Cases available in analysis	Event ^a	128	93,4%
	Censored	9	6,6%
	Total	137	100,0%

Omnibus Tests of Model Coefficients^a

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change ...
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square
1011,760	,008	1	,928	,008	1	,928	,008

Der er absolut ikke signifikant forskel -
men hvad med kvantificering af forskellen?

Rater = hazards = intensiteter

- ▶ Vi har set, at levetiden ikke dur som outcome, fordi en sådan ikke altid observeres.
- ▶ I stedet vil vi gerne modellere “hvor hurtigt overlevelseskurven går ned ad bakke”, altså en mere *dynamisk* beskrivelse, i form af den umiddelbare risiko for død (*rater*, *hazards* eller *intensiteter*):

Givet, at jeg er i live lige nu, hvad er så min risiko for at dø indenfor det næste år?

Eller næste måned? Næste dag/næste time....

- ▶ **Hazard funktion $h(t)$:** den *umiddelbare* risiko for at dø i det næste korte interval Δt er $h(t)\Delta t$

$h(t)$ er en rate, ikke en sandsynlighed



Hazard funktionen

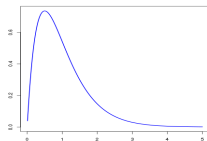
- ▶ er ikke så god at estimere, fordi den hopper op og ned (andel døde i hvert lille tidsrum er enten 0 eller et ret lille tal)
- ▶ estimer i stedet **kumuleret hazard, $H(t)$**
- ▶ Matematikken kan vise, at $H(t) = -\log(S(t))$
- ▶ Hældningen af $H(t)$ afspejler den umiddelbare risiko/rate/hazard $h(t)$ for at dø:
 - ▶ stejl hældning: stor risiko
 - ▶ lille hældning: lav risiko

Teknisk note: $H'(t) = -\frac{S'(t)}{S(t)}$, dvs. det er altså den *relative* ændring af overlevelseskurven, vi modellerer.

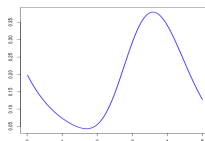


Eksempler på hazard-funktioner $h(t)$

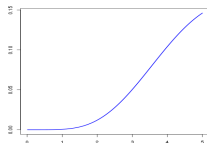
Umiddelbar farlig behandling:



Gavnlig behandling,
med senfølger:

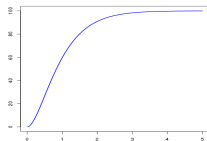


Alder som mest betydende:

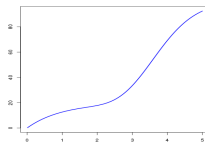


Kumulerede hazard-funktioner $H(t)$

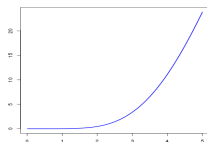
Umiddelbar farlig behandling:



Gavnlig behandling,
med senfølger:

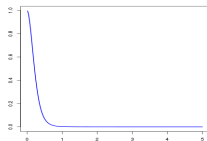


Alder som mest betydende:

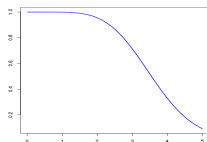


Overlevelseskurver $S(t)$

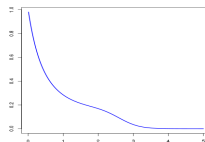
Umiddelbar farlig behandling:



Alder som mest betydende:



Gavnlig behandling,
med senfølger:



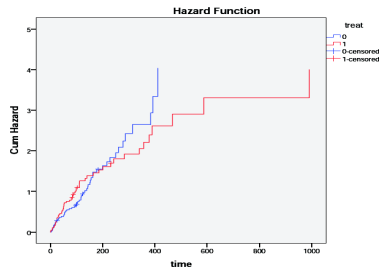
Kumulerede hazard funktioner

i vores eksempel om krigsveteraner

dannes via Analyse/Survival/Kaplan-Meier (se mere s. 85-86)

Under Options
afkrydses Hazard

Er de ens?



Hvis de ikke er ens, er de så **proportionale**? (de starter jo begge i 0)

Cox *proportional hazards* model

En såkaldt **Cox model** specificerer alle hazard-funktioner til at være proportionale.

Her har vi bare to grupper, så vi **antager**

$$h_1(t) = k \times h_0(t)$$

Men **baseline hazard funktionen** $h_0(t)$ specificeres ikke nærmere, dvs. vi antager ikke noget om afhængigheden af tiden

Vi vil gerne estimere proportionalitetskonstanten k ,
hvis modellen altså holder....., dvs. hvis der er additivitet på log-skala: $\log(h_1(t)) = \log(k) + \log(h_0(t))$



Estimat for behandlingseffekt

under antagelse af proportionale intensiteter:

Vi benytter opsætningen som s. 24, idet vi i Options afkrydser CI for $\exp(\beta)$.

Desuden går vi ind i Categorical og trykker på treat, hvorefter vi nede under Reference Category skifter fra Last til First for at få effekten af den nye behandling i forhold til den traditionelle, og ikke omvendt. Husk at trykke Change/Continue.

Hazard ratio for $\text{treat}=1$ vs. $\text{treat}=0$ kommer nu frem under Exp(B) , efterfulgt af konfidensintervallet.



Estimat for behandlingseffekt, II

Parameterestimatet er her differensen mellem grupperne på log-hazard skala, og den skal tilbagetransformeres til en faktor, kaldet **hazard ratio** eller *rate ratio*.

Categorical Variable Codings^a

		Frequency	(1)
treat ^b	0	69	0
	1	68	1

a. Category variable: treat

b. Indicator Parameter Coding

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
treat	,016	,181	,008	1	,928	1,016	,713	1,448

Her er forskellen positiv, så hazard ratio bliver større end 1, nemlig 1.016, med CI=(0.713, 1.448).



Fortolkning af behandlingseffekten

Effekten af eksperimentel behandling ($\text{treat}=1$ vs. $\text{treat}=0$) er at øge hazard med faktoren 1.017, altså en forøgelse på 1.7%.

Dette er dog på ingen måde signifikant ($P=0.93$), idet konfidensintervallet er $CI=(0.71, 1.45)$, og altså indeholder effekter lige fra en forøgelse af hazard med 45% til en formindskelse af hazard med 29%.

Dette er den simplest tænkelige Cox-model med kun en enkelt kategorisk kovariat på 2 niveauer (2 grupper), *og raterne antages altså at være proportionale.*



*Proportionale rater/hazards

I tilfælde af to grupper har vi $h_1(t) = k \times h_0(t)$
og dermed parallellitet på log-skala, også for de *kumulerede* rater:

$$\log(H_1(t)) = \log(k) + \log(H_0(t))$$

Da der gælder $H(t) = -\log(S(t))$, svarer det til, at

$$\log(-\log(S_1(t))) = \log(k) + \log(-\log(S_0(t)))$$

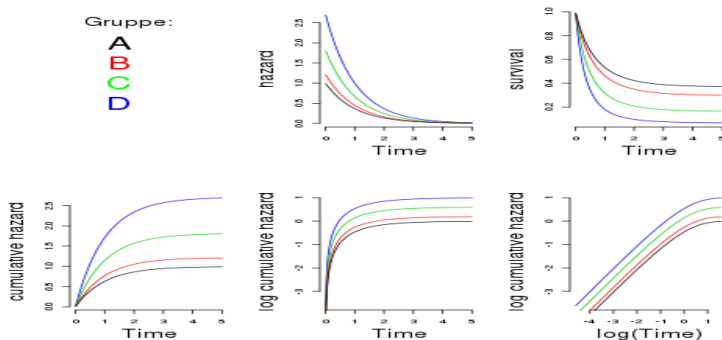
og man siger derfor, at der er parallellitet på loglogs-skala,
eller på *cloglog*-skala. Dette kan kontrolleres

- ▶ Grafisk: se s. 38 og 41
- ▶ Numerisk: se s. 74ff



Eksempel på proportionale rater/hazards

i 4 grupper (4 værdier af en kategorisk kovariat):



Figuren af $\log H(t)$ mod $\log(t)$ nederst til højre viser tydeligt parallelle kurver, dvs. **proportional hazards (PH)**

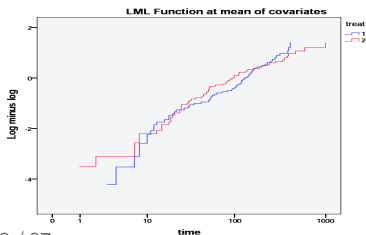
Har vi PH i vores eksempel?

For at undersøge dette, benyttes

Analysis/Survival/Cox Regression, hvor vi sætter

- ▶ time over i feltet Time
- ▶ status i feltet Status, tryk herefter på Define Event og skriv 1 i Single Value, herefter Continue
- ▶ treat over i Strata

Under Plots afkrydser Log minus Log. Efterfølgende kan man redigere grafen til at have logaritmisk tidsakse.



Er disse to parallelle?

Ikke synderligt....

men ret ens...

Check af proportionale rater

Grafisk: Her er to muligheder:

1. Check parallellitet af $H(t)$ 'erne i grupperne, på loglogs-skala, som vist s. 38
2. Benyt de såkaldte *kumulerede score-residualer*, i SPSS kaldet **Partial residuals**

Numerisk: Her er i SPSS vistnok kun en enkelt mulighed:

- Inkluder en tidsafhængig effekt af kovariaten, se mere s. 74ff



*Partielle residualer (Schoenfeld)

Til hvert tidspunkt for event, f.eks. t_i for individ i ser vi på vedkommendes kovariat x_i og spørger:

Er det "den forventede" person, der dør?

$$x_i - \bar{x}(t_i, \hat{\beta})$$

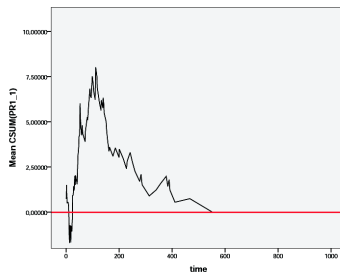
Disse kumuleres over tid og standardiseres (*standardized score process*) og kan afsløre, om betydningen af x er den samme over tid (svarende til proportional hazards):

Hvis f.eks. folk med høje værdier af kovariaten dør "*for tidligt*" i starten, vil processen starte med at være positiv, for derefter at synke ned mod 0.



Grafisk check af proportionale rater

Ved at gemme de partielle residualer fra Cox regressionen (se opsætning s. 24) og selv definere de kumulerede af slagsen, kan det lade sig gøre at producere nedenstående figur (se mere s. 90):



Bemærkninger næste side

Numerisk check af proportionale rater

Figuren s. 41 viser, at det i starten er patienter med høje værdier af variabelen `treat` (dvs. personer med 1-taller for indikatoren for `treat=1`, altså de eksperimentelt behandlede), der dør for tidligt. Dette svarer også til, hvad vi har set på overlevelseskurven (s. 20) og kumuleret hazard funktion (s. 38).

Figuren i SPSS giver dog ingen simulerede forløb, og det er derfor vanskeligt visuelt at bedømme, om den observerede proces kommer for langt væk fra 0-linien.

Det numeriske test kan tilsyneladende heller ikke laves i SPSS - her er man i stedet nødt til at benytte en tidsafhængig effekt af kovariaten, se mere s. 74ff.

Numerisk check af proportionale rater - SAS

Dette er taget fra SAS, jvf. s. 42.

SAS giver 20 simulerede forløb på grafen, og sammenholdt med disse kan man se, at vores proces ud til at være ret aparte, så antagelsen om proportional hazards er ikke helt god.

Dette fremgår også af det dertil hørende test, baseret på 1000 simulationer:

Supremum Test for Proportionals Hazards Assumption

Variable	Maximum Absolute Value	Replications	Seed	Pr > MaxAbsVal
treat1	1.4529	1000	106165	0.0240



Antagelsen om proportionale intensiteter er ikke opfyldt

Det betyder at:

- ▶ Forskellen på behandlingerne er *ikke* er konstant over tid.
- ▶ Hazard-funktionerne har ikke samme form for de to grupper

Hvad gør vi så?

- ▶ Overvejer, om vi har glemt nogle vigtige kovariater:
Fik vi svaret på det rigtige?
- ▶ Benytter tidsafhængige kovariater
Dette er svært, så vi udskyder det til s. 74ff



Har vi andre kovariater?

Jada, vi har adskillige:

celltype: 1:squamous, 2:smallcell, 3:adeno, 4:other

karno: Karnofsky performance score (0-100, 100 er godt):

- ▶ 10-30: hospitaliseret
- ▶ 40-60: delvist hospitaliseret
- ▶ 70-90: tager vare på sig selv

diagtime: Måneder fra diagnose til randomisering/behandling

age: Personens alder i år

prior: Er der givet tidligere behandling? 0:nej, 1:ja

Vi forsøger os med en model, der (ud over `treat`) tager højde for `celltype` og `karno`, dvs. en model med to kategoriske og en enkelt kvantitativ kovariat.



Regressionsmodeller i overlevelsesanalyse

i daglig tale: **Cox modeller:**

Da hazard-funktioner er positive, $h(t) > 0$, vil $\log(h(t))$ kunne antage alle værdier, og det er derfor naturligt at modellere effekten af kvantitative kovariater (x , f.eks. karno) på denne skala:

$$\log(h(t|x)) = \log(h_0(t)) + \beta x + \dots \text{flere}$$

således at effekten bliver *multiplikativ* på selve hazard-skalaen.

Her er $h_0(t)$ **baseline hazard**, svarende til $x = 0$

(altså en generaliseret form for intercept),

og β er regressionskoefficienten hørende til kovariaten x .

Bemærk, at β er uafhængig af tiden,

og at $\exp(\beta)$ netop er proportionalitetskonstanten k fra s. 32



Hazard-funktioner

For hver kombination af kovariater har vi til hvert tidspunkt en vis umiddelbar rate for død, som har det multiplikative udtryk:

$$h(t|x) = h_0(t) \cdot \exp(\beta x)$$

Hazard funktionen (raten) er altså en funktion af to (eller flere) variable: tiden og kovariaten (kovariaterne):

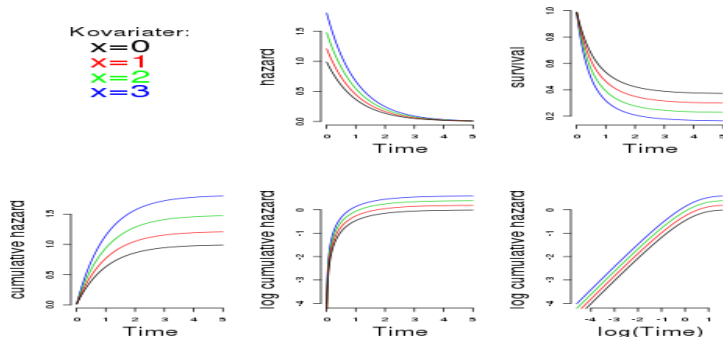
Proportional hazards: er en antagelse om afhængigheden af tiden, nemlig at denne afhængighed er den samme for alle kovariat-værdier, pånær en multiplikativ effekt (en faktor)

Linearitet: er en antagelse om afhængigheden af hver af de kvantitative kovariater, nemlig at der for hver givet tidspunkt er linearitet (på log-skala)



Eksempel på hazard-funktioner

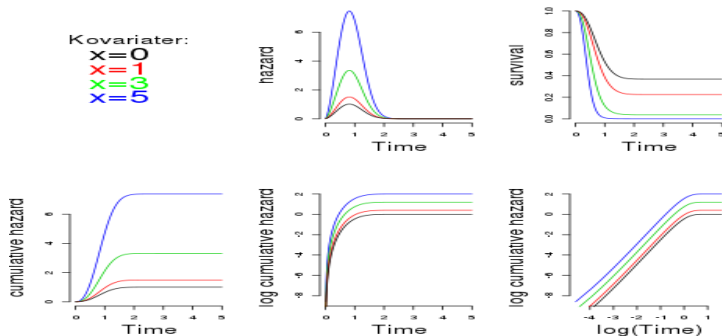
for 4 forskellige ækvidistante værdier (0, 1, 2 og 3) af en kvantitativ kovariat x :



Her er både **proportional hazards** og **linearitet** (det første ses ved parallellitet, og det sidste fordi afstanden mellem kurverne er ækvidistant)

Eksempel på hazard-funktioner

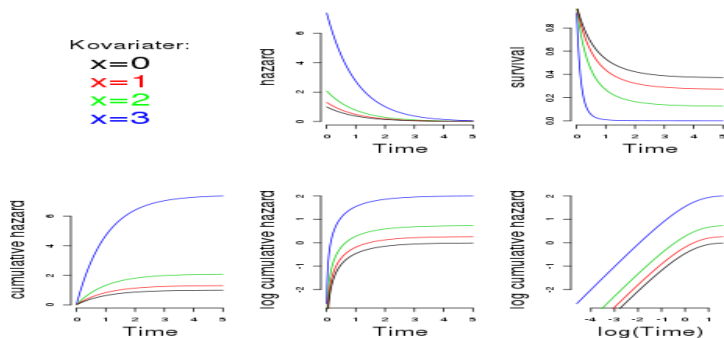
for 4 forskellige værdier af en kvantitativ kovariat x :



Her er både **proportional hazards** og **linearitet** (det første ses ved parallellitet, og det sidste fordi afstanden mellem kurverne afspejler de ikke-ækvivalente kovariat-værdier)

Eksempel på hazard-funktioner

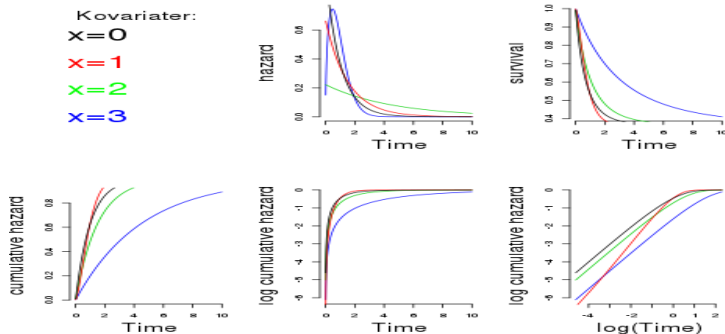
for 4 forskellige ækvidistante værdier (0, 1, 2 og 3) af en kvantitativ kovariat x :



Her er **proportional hazards**, men **ikke** linearitet (kurverne er parallelle, men **ikke** ækvidistante)

Eksempel på hazard-funktioner

for 4 forskellige ækvidistante værdier (0, 1, 2 og 3) af en kvantitativ kovariat x :



Her er *hverken* proportional hazards, *eller* linearitet

Model med 3 kovariater

- ▶ `treat`: Behandlingen
- ▶ `celltype`: De 4 cancertyper
- ▶ `karno`: kvantitativ Karnofsky score

Vi benytter `Analyse/Survival/Cox Regression`, hvor vi sætter

- ▶ `time over` i feltet `Time`
- ▶ `status` i feltet `Status` (med 1 som event)
- ▶ såvel `treat` som `celltype` og `karno` sættes i `Covariates`

Herefter går man ind i `Categorical` og sørg for, at både `treat` og `celltype` står under `Categorical Covariates`.

Her kan man også ændre reference-værdier for disse, hvis det ønskes, se s. 88

Sørg for, at der under `Options` er afkrydset i `CI` for $\exp(\beta)$



Output fra model med 3 kovariater

Modellen s. 52 giver outputtet:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
treat	,257	,201	1,645	1	,200	1,293	,873	1,917
celltype			17,301	3	,001			
celltype(1)	1,148	,295	15,142	1	,000	3,151	1,768	5,617
celltype(2)	,393	,282	1,939	1	,164	1,481	,852	2,576
celltype(3)	,820	,269	9,297	1	,002	2,270	1,340	3,844
karno	-,031	,005	36,260	1	,000	,969	,960	,979

Betegnelserne for de 4 typer er:

1:adeno, 2:large, 3:smallcell, 4:squamous

Se fortolkning s. 54-55



Fortolkning af output

- ▶ Der er stærkt signifikant negativ effekt af karno-score ($P = 0.000$)
- ▶ Der er stærkt signifikant forskel på cancertyperne, med adeno og smallcell som de farligste (de har de største positive afvigelser fra referencegruppen squamous)
- ▶ Behandlingseffekten er fortsat ikke signifikant, $P=0.20$

For at forstå selve estimerne, skal vi se på rate ratioerne (hazard ratios) fra s. 53.



Fortolkning af rate ratioer

- ▶ Effekten af 1 enhed på karno-skalaen er et fald i dødsraten på 3.1%, $CI=(2.1\%, 4.0\%)$, se dog også næste side, hvor vi får effekten af 10 enheder ved at definere en ny variabel $karno10=karno/10$, således at 1 enhed af $karno10$ svarer til 10 enheder af karno.
- ▶ Umiddelbart er den eksperimentelle behandling dårlig, med en forøget rate på 29.3%, men CI viser, at det ligeså godt kunne være en formindsket rate (12.7% lavere), eller muligvis helt op til næsten dobbelt så stor rate (91.7% højere)
- ▶ adeno-typen estimeres til at være ca. 3 gange så "farlig" som referencen squamous, dog med betydelig usikkerhed, svarende til faktorer mellem 1.77 og 5.62.



Output fra model med 3 kovariater, fortsat

I SPSS må vi regne videre på estimerterne fra modellen s. 52-53 ved *håndregning*, eller ved at bruge en ny kovariat som f.eks. karno10, defineret s. 55

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
treat	-,257	,201	1,645	1	,200	,773	,522	1,146
celltype			17,301	3	,001			
celltype(1)	1,148	,295	15,142	1	,000	3,151	1,768	5,617
celltype(2)	,393	,282	1,939	1	,164	1,481	,852	2,576
celltype(3)	,820	,269	9,297	1	,002	2,270	1,340	3,844
karno10	-,311	,052	36,260	1	,000	,733	,662	,811

Fortolkning:

Effekten af en øgning på 10 enheder på karno-skalaen er et fald i dødsraten på 26.7%, CI=(18.9%, 33.8%)



Output fra model med 3 kovariater, fortsat

I SPSS har man ikke umiddelbart muligheden for at sammenligne de 4 cancertyper parvist. Man får udelukkende sammenligninger til referencegruppen.

Hvis man vil se andre specifikke sammenligninger, må man enten:

- ▶ benytte *håndregning*:
det kan godt lade sig gøre at udregne alle de parvise estimer, men man kan ikke få konfidensgrænser på uden programmets hjælp.
- ▶ omdefinere grupperne, f.eks. omdøbe adeno til 4:adeno og de øvrige til noget med 1,2 og 3. Så kommer 4:adeno sidst i “alfabetet” og vil blive brugt som referencegruppe.



Modelcheck

For kategoriske kovariater:

Her drejer det sig om `treat` og `celltype`:

Her skal antagelsen om *Proportional hazards* checkes:

- ▶ ved at sætte kovariaten i `strata` i stedet for i `Covariates`, for derefter at tegne $\log H(t)$ (afkryds `Log minus log under Plots`), se det tidligere check af `treat` s. 38, og det nuværende på s. 59

For kovariaten `celltype` finder vi her figuren på s. 60

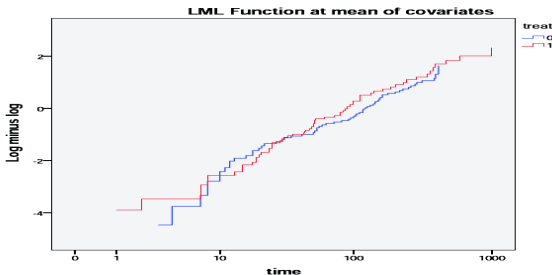
- ▶ benytte kumulerede partielle residualer og tegne disse op mod tiden for se, om der er et udpræget mønster (desværre uden simulationer og test som i SAS og R), jvf. s. 40-43.



Modelkontroltegninger for treat

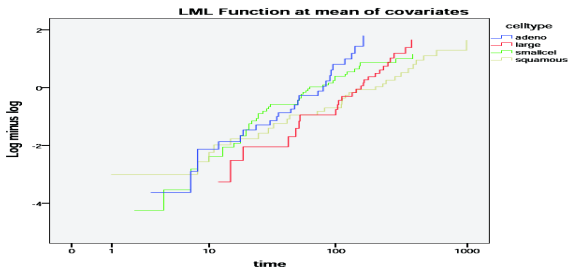
i denne nye model med 3 kovariater.

Da treat er en kategorisk variabel, sætter vi den i stedet i strata-statement for at kunne tegne kumulerede rater, på log-skala, se fremgangsmåde for tilsvarende figur på s. 89.



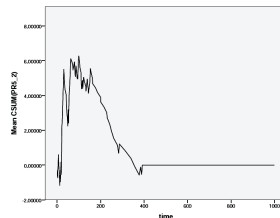
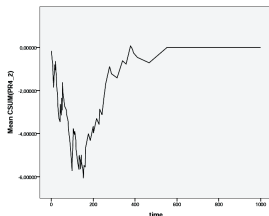
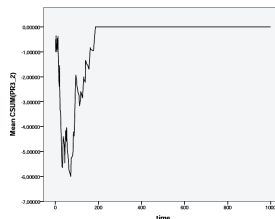
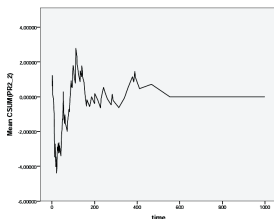
Modelkontroltegninger for celltype

Til brug for denne figur, er det `celltype`, der sættes i `strata`-statement for at kunne tegne kumulerede rater, på log-skala, se tilsvarende fremgangsmåde s. 89



Det ses, at det netop er referencegruppen `squamous`, der opfører sig anderledes end de andre, med lavere hældning

Proportional hazards via partielle residualer



Bemærkning vedrørende Proportional hazards

Numeriske tests af modeltilpasning er desværre ikke muligt i SPSS, med mindre man benytter tidsafhængig effekt af kovariaten, se mere s. 74ff

Fra SAS får vi nogle numeriske tests, baseret på simulationer af de kumulerede partielle residualer:

Først for kovariaten `treat`:

Det numeriske test giver nu $P=0.42$, dvs.

vi har ikke mere anledning til at betvivle proportionaliteten

Bemærk, at vi tidligere (s. 43) havde $P=0.024$, og de proportionale intensiteter er altså opnået ved at tilføje vigtige ekstra kovariater



Bemærkning vedrørende Proportional hazards, II

Og så for kovariaten `celltype`:

Det numeriske test er her 3-delt, idet alle typer sammenlignes til reference-typen `squamous`. P-værdierne er:

`adeno`: 0.036, `large`: 0.024 og `smallcell`: 0.050,

så alle er signifikante (den ene dog lige på 0.05..., men alligevel).

Her har vi altså et problem, og fra figuren s. 60 ses det, at netop celletypen `squamous` (den olivengrønne, svær at se) adskiller sig fra de øvrige ved at have et fladere forløb, svarende til en større dødelighed fra start, efterfulgt af lavere dødelighed senere.

Havde vi valgt en anden referencegruppe, havde vi rimeligvis kun set problemer i forhold til sammenligningen med `squamous`.



Modelcheck for kvantitative kovariater

Her drejer det sig kun om karno:

Her skal der checkes to antagelser:

1. *Linearitet* på $\log(\text{hazard})$ -skala: Her kan man benytte
Kumulerede Martingal residualer:

Disse kan vist ikke fås i SPSS??

For hver kovariatværdi x ser man på, i hvor høj grad personer med kovariatværdier $< x$ "levede for længe" (positive værdier) eller "døde for tidligt" (negative værdier).

Metoden er dog ikke særligt sensitiv

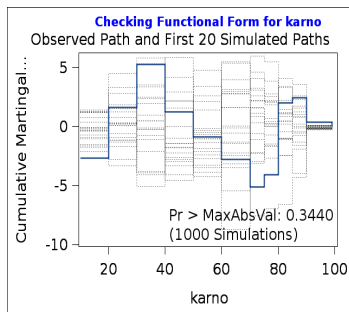
2. *Proportional hazards*: checkes (næsten) som for kategoriske kovariater, nemlig via de kumulerede partielle residualer.



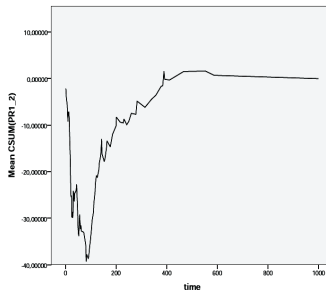
Check af modelforudsætninger for karno

Lineariteten:

Dette er en SAS-figur,
se s. 64



Proportional hazards:



Linearitetsantagelsen for karno

- ▶ Figuren til venstre s. 65 og det tilhørende test (fra SAS) viser, at lineariteten er rimelig ($P=0.34$).
- ▶ Alternative checks af lineariteten kunne være:
 - ▶ Såvel karno som $\log(\text{karno})$ som kovariater (s. 67)
 - ▶ En lineær spline i karno, hvor vi f.eks. tillader knæk ved $\text{karno}=40$ og $\text{karno}=80$:

Se mere s. 93, samt dele af output s. 68



Ekstra kovariat afledt af karno

nemlig $\log karno = \log_2(karno)$

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
karno	-,034	,016	4,265	1	,039	,967	,936	,998
treat	,262	,203	1,675	1	,196	1,300	,874	1,934
celltype			17,277	3	,001			
celltype(1)	1,157	,300	14,904	1	,000	3,181	1,768	5,724
celltype(2)	,395	,282	1,953	1	,162	1,484	,853	2,581
celltype(3)	,817	,269	9,194	1	,002	2,263	1,335	3,836
logkarno	,084	,493	,029	1	,864	1,088	,414	2,862

Da $\log karno$ ikke er tilnærmelsesvist signifikant, er der ingen tegn på problemer med linearitetsantagelsen.

Effekt af karno som lineær spline

Vi definerer nu karno40 således:

`COMPUTE karno40=(karno>40)*(karno-40)`

og karno80 på tilsvarende vis.

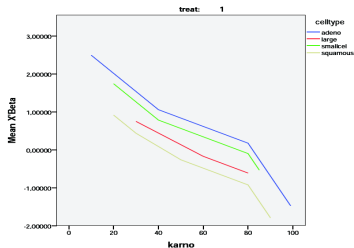
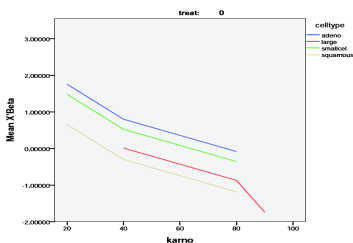
Disse indsættes nu (sammen med karno, hvorved vi får:

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI for Exp(B)	
							Lower	Upper
karno	-,048	,016	8,694	1	,003	,953	,923	,984
treat	,259	,202	1,646	1	,200	1,296	,872	1,926
celltype			16,943	3	,001			
celltype(1)	1,100	,298	13,648	1	,000	3,005	1,676	5,387
celltype(2)	,312	,286	1,195	1	,274	1,366	,781	2,391
celltype(3)	,826	,268	9,471	1	,002	2,285	1,350	3,867
karno40	,026	,021	1,531	1	,216	1,026	,985	1,069
karno80	-,065	,049	1,733	1	,188	,937	,851	1,032

Effekt af karno som lineær spline

I analysen kan man gemme predikterede værdier ved at afkrydse $X \cdot \text{Beta}$ under Save. Disse kommer til at hedde XBE_1, og de kan nu benyttes til at tegne illustrere effekten af karno:



Test af knæk for effekt af karno

Den lineære spline tillader knæk ved karno=40 og karno=80,
men er de nødvendige?

Det kan man få svar på ved at køre to modeller på en gang (se s. 94), en med de to kovariater karno40 og karno80 og en uden. Metoden giver her testene

Variables not in the Equation^a

	Score	df	Sig.
karno40	,781	1	,377
karno80	1,117	1	,291

a. Residual Chi Square = 2,628 with 2 df Sig. = ,269

og dermed ingen evidens for problemer med lineariteten i karno ($P=0.27$).



Check af proportional hazards for karno

Testet (fra SAS) for proportionale intensiteter for karno (s. 65) viser, at denne antagelse er helt gal ($P < 0.0001$).

Denne manglende proportionalitet afspejles i figuren til højre på s. 65, hvor de kumulerede partielle residualer ses at have et tydeligt mønster.

Dette betyder:

- ▶ Effekten af karno er *ikke* konstant over tid.
- ▶ Hazard-funktionerne har ikke samme form for alle værdier af karno

Hvad gør vi så??



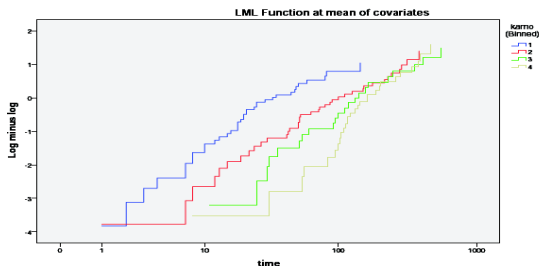
Vi prøver at forstå problemet

- ▶ På figuren s. 65 ses, at de med små værdier af karno i højere grad dør i starten af perioden i forhold til, hvad modellen siger
- ▶ Effekten af karno er i sig selv negativ, således at små værdier af karno altid vil formodes at dø først - de gør det bare endnu mere i starten
- ▶ Det betyder, at effekten af karno er størst i starten, men aftager efterhånden som tiden går



Hvis vi nu grupperer karno-kovariaten

og sætter den ind i strata:



Her ser vi ligeledes, at kurverne afviger en del i starten (svarende til stor effekt af karno), men efterhånden nærmer sig hinanden (svarende til *næsten ingen* effekt)

Tidsafhængig effekt af kovariater

Vi har altså set, at antagelsen om proportionale intensiteter (**hazards**) ikke holdt for kovariaten karno.

Det betyder, at den prognostiske værdi af karno ændrer sig med tiden, og vi fortolkede det som at effekten af karno er stærkest i starten.

Hvad gør vi så?

Vi kan opdele tiden i intervaller, og estimere en proportionalitetsfaktor for hvert interval for sig.

Vi kan f.eks. opdele tiden efter **kvartilerne for de døde**:
23, 62 og 147.5 dage



*Tidsafhængige kovariater i praksis

- ▶ For at få effekten af 10 karno-enheder direkte, skifter vi karno-variablen ud med $\text{karno10} = \text{karno}/10$, således at 1 enhed af karno10 svarer til 10 enheder af karno.
- ▶ Dernæst indfører vi separate effekter af karno10 ved direkte i Cox-analysen at definere en slags *interne kovariater* i de forskellige tidsrum. Dette kan kun gøres i syntax'en

```
TIME PROGRAM.  
COMPUTE karno1 = karno10*(t_ gt 0 and t_ le 23).  
COMPUTE karno2= karno10*(t_ gt 23 and t_ le 62).  
COMPUTE karno3 = karno10*(t_ gt 62 and t_ le 147.5).  
COMPUTE karno4 = karno10*(t_ gt 147.5).  
COXREG time  
  /STATUS=status(1) /CONTRAST(treat)=Indicator(1) /CONTRAST(celltype)=Indicator  
  /METHOD=ENTER celltype treat karno1 karno2 karno3 karno4  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

Disse 4 “karno-variable skal nu erstatte den oprindelige karno, hvorved vi får outputtet s. 76



Output fra tidsafhængig effekt af karno

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
celltype			16,463	3	,001	
celltype(1)	1,091	,300	13,198	1	,000	2,978
celltype(2)	,334	,284	1,386	1	,239	1,397
celltype(3)	,880	,277	10,119	1	,001	2,412
treat	,083	,205	,165	1	,685	1,087
karno1	-,514	,093	30,685	1	,000	,598
karno2	-,364	,091	15,957	1	,000	,695
karno3	-,077	,122	,401	1	,527	,926
karno4	-,032	,128	,064	1	,801	,968

Se fortolkning næste side



Fortolkning af tidsafhængig effekt

Effekten af karno aftager med tiden:

Rate-ratioerne nærmer sig 1, ganske som vi fandt det tidligere på kontroltegningen s. 73.

Det kunne have den simple forklaring, at karno var målt ved indgang i studiet.

Ved at definere to modeller på samme tid (se mere s. 95), kan vi desuden få et test af tidsafhængigheden:

Variables not in the Equation^a

	Score	df	Sig.
karno1	7,379	1	,007
karno2	,456	1	,500
karno3	3,938	1	,047

a. Residual Chi Square = 13,168 with 3 df Sig. = ,004



Kan der være interaktioner med celltype?

Man kunne forestille sig en differentieret effekt af behandlingen for de 4 celletyper, altså interaktionen `treat*celltype` (se kode s. 97):

Variables in the Equation						
	B	SE	Wald	df	Sig.	Exp(B)
celltype			8,086	3	,044	
celltype(1)	,970	,456	4,520	1	,034	2,638
celltype(2)	-,195	,397	,242	1	,623	,823
celltype(3)	,373	,339	1,213	1	,271	1,453
treat	-,603	,406	2,198	1	,138	,547
karno1	-,508	,094	29,129	1	,000	,602
karno2	-,367	,093	15,459	1	,000	,693
karno3	-,076	,126	,362	1	,548	,927
karno4	-,019	,134	,021	1	,886	,981
celltype*treat			6,142	3	,105	
celltype(1)*treat	,367	,600	,375	1	,541	1,444
celltype(2)*treat	,989	,570	3,008	1	,083	2,688
celltype(3)*treat	1,151	,518	4,939	1	,026	3,161

Der er ikke tilstrækkelig evidens for en interaktion ($P=0.105$)

Interaktioner med celltype, fortsat

Estimerer for behandlingen, opdelt efter celletype:
problemer med SPSS-så dette er fra SAS

Analysis of Maximum Likelihood Estimates

Parameter			DF	Parameter Estimate	Standard Error	Chi-Square
celltype	adeno		1	0.96993	0.45621	4.5201
celltype	large		1	-0.19513	0.39677	0.2419
celltype	smallcel		1	0.37338	0.33907	1.2126
celltype	squamous		0	0	.	.
treat*celltype	1	adeno	1	-0.23554	0.42944	0.3008
treat*celltype	1	large	1	0.38605	0.40013	0.9309
treat*celltype	1	smallcel	1	0.54845	0.32845	2.7882
treat*celltype	1	squamous	1	-0.60258	0.40650	2.1974
treat*celltype	0	adeno	0	0	.	.
treat*celltype	0	large	0	0	.	.
treat*celltype	0	smallcel	0	0	.	.
treat*celltype	0	squamous	0	0	.	.
karno1			1	-0.50791	0.09411	29.1285
karno2			1	-0.36698	0.09334	15.4592
karno3			1	-0.07567	0.12582	0.3617
karno4			1	-0.01926	0.13426	0.0206



Problemstillinger, I: Forsinket indgang

Forsinket indgang = left-truncation

Når vi skal sammenligne behandlinger, kan patienterne naturligvis først tælle med, når de blev randomiseret til en af disse og derfor er en del af *risikogruppen*.

Hvis vi benytter en tidsskala, der hedder *tid siden diagnose*, kan vi have et problem, idet nogle kan være døde efter diagnose og inden randomisering, således at de, der indgår i studiet er selekterede, måske forskelligt i de to behandlingsgrupper. **Hvad gør vi så?**

Dette kan (måske) gøres i SPSS ved at benytte Analyze/Complex Samples/Prepare for Analysis, efterfulgt af Analyze/Complex Samples/Cox Regression, hvor man på passende sted skal indsætte en variabel, der indeholder "*time to entry*"



Problemstillinger, II: Competing risks

Competing risks = konkurrerende dødsårsager/afgangsårsager:

Hvis man f.eks. studerer tid til tilbagefald efter behandling, har man det problem, at folk stadig kan dø, *inden* de får tilbagefald. **Hvad gør man så?**

1. Man analyserer det sammensatte endepunkt: tid til *enten* tilbagefald *eller* død, hvad der så end kommer først
2. Man opfatter død som censureringer

I første tilfælde er der intet problem, men i tilfælde 2 får man problemer:

Man kan *stadig* godt

- Udregne hazard-funktioner, og estimere i Cox-modeller for tilbagefald=recidiv, ved blot at opfatte død som en censurering, men...



Competing risks, fortsat

men man kan *ikke*

- ▶ fortolke overlevelseskurver (Kaplan-Meier), se s. 20 fordi det ikke giver nogen mening at antage, at de censurerede patienter (som jo er døde) ville få recidiv i samme takt som de ucensurerede, se s. 15.
- ▶ beregne kumulerede incidenser, altså sandsynligheder for at opleve recidiv (som vist s. 21)

I stedet kan man evt. benytte de såkaldte **Fine and Gray modeller**, der direkte modellerer kumulative incidenser.

Pas på: Hvis en behandling slår folk ihjel hurtigt, når de jo ikke at få recidiv!



Ofte stillede spørgsmål: FAQ

- ▶ *Hvorfor bruger vi ikke bare logistisk regression af f.eks. 1-års mortaliteten?*

Fordi vi mister en masse information, og fordi vi mister muligheden for at håndtere censureringer

- ▶ *Hvorfor bruger vi ikke bare normalfordelingsmodeller på tiden fra randomisering til død?*

Fordi fordelingen ofte er ganske anderledes end en normalfordeling, og igen fordi vi mister muligheden for at håndtere censureringer

Overlevelsesanalyse er den stærkeste måde at håndtere overlevelsesdata på!



APPENDIX

med SPSS-vejledning svarende til nogle af slides

- ▶ Kaplan-Meier kurver, s. 85
- ▶ Kumulerede incidenser, s. 86
- ▶ Kumulerede hazards, s. 86
- ▶ Log-rank test, s. 87
- ▶ Cox regression, s. 88
- ▶ Check af proportional hazards, s. 89-90
- ▶ Model med 3 kovariater og modelkontrol, s. 91-93
- ▶ Tidsafhængige kovariater, s. 95
- ▶ Interaktioner, s. 97



Kaplan-Meier kurver i SPSS

Slide 20

Vi benytter Analyse/Survival/Kaplan-Meier, hvor vi sætter

- ▶ time over i feltet Time
- ▶ status i feltet Status, tryk herefter på Event og skrive 1 i Single Value, herefter Continue
- ▶ treat i Factor

Gå derefter ind i Options, og sørg for, at der er afkrydset i Survival.

Hvis man vil have konfidensgrænser på, er man nødt til at gøre et stort stykke arbejde selv, men det er muligt ved hjælp af Save-knappen at gemme standard error på disse overlevelseskurver



Kumulerede incidenser og kumulerede hazards

Slide 21, 31

Følg opsætningen fra s. 85, men sørg for, at der under Options nu er afkrydset i

- ▶ One minus survival for at få de kumulerede incidenser (s. 21)
- ▶ Hazard: for at få de kumulerede hazards (s. 31)



Log-Rank test

Slide 24

Vi benytter Analyse/Survival/Cox Regression, hvor vi sætter

- ▶ time over i feltet Time
- ▶ status i feltet Status, tryk herefter på Event og skriv 1 i Single Value, herefter Continue
- ▶ treat over i Covariates (eller Block 1 of 1)

Log-rank testet aflæses nu under Overall(score)



Cox regression

Slide 32ff

Vi benytter opsætningen som s. 87, idet vi i `Options` afkrydser `CI` for $\exp(\beta)$.

Desuden går vi ind i `Categorical` og trykker på `treat`, hvorefter vi nede under `Reference Category` skifter fra `Last` til `First` for at få effekten af den nye behandling i forhold til den traditionelle, og ikke omvendt. Husk at trykke `Change/Continue`.

Hazard ratio for `treat=1` vs. `treat=0` kommer nu frem under `Exp(B)`, efterfuldt af konfidensintervallet.



Check af proportionale rater/hazards, figur

Slide 38

Vi benytter Analyse/Survival/Cox Regression, hvor vi sætter

- ▶ time over i feltet Time
- ▶ status i feltet Status, tryk herefter på Event og skrive 1 i Single Value, herefter Continue
- ▶ treat over i Strata

Under Plots afkrydser Log minus Log

Dette giver dog ikke en logaritmisk tidsakse, men **det kan man efterfølgende bede om under redigering af grafen.**



Check af proportionale rater/hazards

Slide 41-42

I Cox-regressionen fra s. 88 benyttes `Save`, hvor der afkrydses i `Partial residuals`. Disse kommer til at hedde `PR1_1`.

For at kumulere disse over tid, må vi først sortere i tids-rækkefølge ved at gå ind i `Data/Sort Cases` og sætte `time` i `Sort by`.

Herefter benyttes `Transform/Compute`, hvorefter man i **Syntax'en** definerer `CUMRES=CSUM(PR1_1)` ved at skrive

```
CREATE CUMRES=CSUM(PR1_1)
```

Variablen `CUMRES` kan nu tegnes op imod `time`



Model med 3 kovariater

Slide 52

Vi benytter Analyse/Survival/Cox Regression, hvor vi sætter

- ▶ time over i feltet Time
- ▶ status i feltet Status (med 1 som event)
- ▶ såvel treat som celltype og karno sættes i Covariates

Herefter går man ind i Categorical og sørg for, at både treat og celltype står under Categorical Covariates.

Her kan man også ændre reference-værdier for disse, hvis det ønskes, se s. 88

Sørg for, at der under Options er afkrydset i CI for $\exp(\beta)$



Model med 3 kovariater, fortsat

Slide 52

Hvis man vil have effekten af 10 enheder på karno-skalaen i stedet for effekten af en enkelt enhed, kan man med fordel benytte variabelen $\text{karno10} = \text{karno}/10$ i stedet for karno

For at sammenligne de enkelte cancerformer parvist, må man omdefinere grupperne successivt for at ændre referencegruppe.



Kovariat modelleret som spline

Slide 66-69

Definitionerne af de to ekstra kovariater (defineres i Transform/Compute) er

```
karno40=(karno>40)*(karno-40)
```

```
karno80=(karno>80)*(karno-80)
```

Vi skal endvidere

- ▶ Gemme de predikterede værdier ($X \cdot \text{Beta}$ under Save), og efterfølgende benytte disse XBE_1 til at illustrere den estimerede effekt af karno, altså den knækkede linie, for hver af de indgående grupper.
- ▶ Definere to modeller på samme tid (se s. 94), så vi kan teste, om begge de to ekstra led kan undværes samtidig, altså om effekten af karno kan beskrives ved simpel linearitet



Kovariat modelleret som spline

Slide 70

Når Cox-modeller sættes op, er der mulighed for at definere flere på samme tid, og disse køres så enkeltvis, men sammenlignes desuden med hinanden.

Fremgangsmåden er:

Efter at have sat Cox-modellen op, klikkes Next, hvor man nu kan tilføje en eller flere kovariater (f.eks. her karno40 og karno80). Herved fås et samlet test for udeladelse af de to tilføjede kovariater.



Tidsafhængige effekt af kovariater

Slide 74ff, (75ff)

SPSS har en “*intern*” tid, kaldet T_{-} , som f.eks. kan benyttes for at definere effekter, der kun skal virke i bestemte intervaller.

Hvis modellen skal indeholde flere af disse samtidig, er man nødt til at definere den i Syntax. Vi skal her bruge disse:

- ▶ $\text{karno1} = \text{karno10} * (T_{-} \leq 23)$
- ▶ $\text{karno2} = \text{karno10} * (23 < T_{-} \leq 62)$
- ▶ $\text{karno3} = \text{karno10} * (62 < T_{-} \leq 147.5)$
- ▶ $\text{karno4} = \text{karno10} * (147.5 < T_{-})$

Disse 4 “karno-variable skal nu erstatte den oprindelige karno (se syntax s. 96), hvorved vi får outputtet s. 76



Tidsafhængige effekt af kovariater, II

Det er kun en ganske speciel form for disse, der kan defineres i menuerne.

Her er man nødt til at benytte syntax:

```
TIME PROGRAM.
```

```
COMPUTE karno1 = karno10*(t_ gt 0 and t_ le 23).
```

```
COMPUTE karno2= karno10*(t_ gt 23 and t_ le 62).
```

```
COMPUTE karno3 = karno10*(t_ gt 62 and t_ le 147.5).
```

```
COMPUTE karno4 = karno10*(t_ gt 147.5).
```

```
COXREG time
```

```
  /STATUS=status(1) /CONTRAST(treat)=Indicator(1) /CONTRAST(celltype)=Indicator
```

```
  /METHOD=ENTER celltype treat karno1 karno2 karno3 karno4
```

```
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```



Interaktion

i model med tidsafhængige kovariater.

Slide 78

Dette fungerer helt som sædvanligt, idet man blot indsætter interaktionsleddet, her `treat*celltype`

```
TIME PROGRAM.  
COMPUTE karno1 = karno10*(t_ gt 0 and t_ le 23).  
COMPUTE karno2= karno10*(t_ gt 23 and t_ le 62).  
COMPUTE karno3 = karno10*(t_ gt 62 and t_ le 147.5).  
COMPUTE karno4 = karno10*(t_ gt 147.5).  
COXREG time  
  /STATUS=status(1) /CONTRAST(treat)=Indicator(1) /CONTRAST(celltype)=Indicator  
  /METHOD=ENTER celltype treat karno1 karno2 karno3 karno4 treat*celltype  
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20).
```

Af en eller anden grund ser det ikke ud til, at man her kan få de enkelte effekter af `treat` frem ved at udelade leddet `treat....??`

