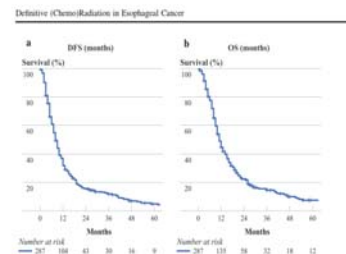


De drie mooiste modellen deel 4

Medische statistiek Survival analyse

Hans Burgerhof
h.burgerhof@home.nl

Uit Smit et al.

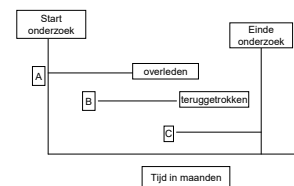


Breastcancer.sav (n =1207)

Name	Type	Width	Decimals	Label	Values	Missing	Columns
1 id	Numeric	8	0		None	None	8
2 age	Numeric	8	0	Age (years)	None	None	8
3 pathsize	Numeric	8	2	Pathologic Tumor	99.00	None	8
4 ipos	Numeric	8	0	Positive Axillary	None	None	8
5 stage	Numeric	8	0	Histologic Grade (4, Unknown)	4	None	8
6 er	Numeric	8	0	Estrogen Receptor (0, Negative)	2	None	8
7 pr	Numeric	8	0	Progesterone R (0, Negative)	2	None	8
8 status	Numeric	8	0	Status	9	None	8
9 pathsize	Numeric	8	0	Pathologic Tumor	99	None	8
10 time	Numeric	8	2	Time (months)	None	None	8

Overleving (beschrijving)
Verschillen tussen groepen?
Model voor survival?

Drie dames, elk 15 maanden in het onderzoek



Censurering !

Balans experiment maximaal 5 minuten



time	status	group
1	0	1
2	0	1
3	0	1
4	0	1
5	0	1
6	0	1
7	0	1
8	0	1
9	0	1
10	0	1
11	0	1
12	0	1
13	0	1
14	0	1
15	0	1
16	0	1
17	0	1
18	0	1
19	0	1
20	0	1
21	0	1
22	0	1
23	0	1
24	0	1
25	0	1

De "survival" tijd
in seconden

controlegroep

met alcohol

0 = niet gevallen
1 = gevallen

Is er verschil in "survival"
tussen de groepen?

Probleem van censurering

- In beide groepen zijn personen gecensureerd.
- Waarom zijn twee mensen voortijdig gecensureerd? Hoelang hadden de mensen die na 5 minuten zijn gecensureerd nog door kunnen gaan?
- Heeft het vergelijken van gemiddelde survivaaltijden zin?
- Een overlevingscurve is een grafiek van *overlevingskansen* die stapsgewijs berekend worden aan de hand van het aantal mensen "at risk"

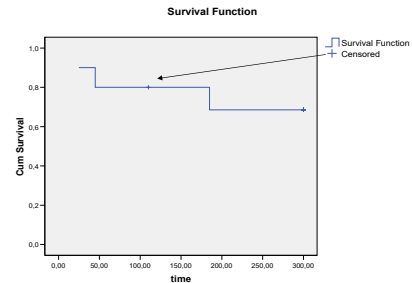
Overlevingskansen (balansexperiment)

$P = 8/9 \cdot 0,9$

	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	25,000	1,00	,900	,095	1	9
2	45,000	1,00	,800	,126	2	8
3	110,000	,00	-	-	2	7
4	185,000	1,00	,686	,151	3	6
5	300,000	,00	-	-	3	5
6	300,000	,00	-	-	3	4
7	300,000	,00	-	-	3	3
8	300,000	,00	-	-	3	2
9	300,000	,00	-	-	3	1
10	300,000	,00	-	-	3	0

$P = 6/7 \cdot 0,8$

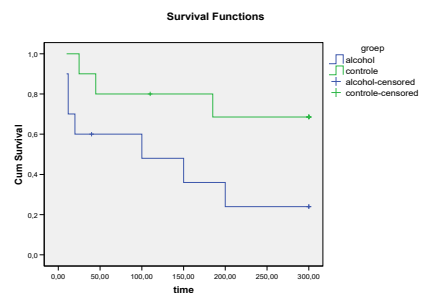
Kaplan-Meier curve (balans data)



Beide groepen

groep	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
controle	1	25,000	1,00	,900	,095	1
	2	45,000	1,00	,800	,126	2
	3	110,000	,00	-	-	2
	4	185,000	1,00	,686	,151	3
	5	300,000	,00	-	-	3
	6	300,000	,00	-	-	3
	7	300,000	,00	-	-	3
	8	300,000	,00	-	-	3
	9	300,000	,00	-	-	3
	10	300,000	,00	-	-	3
alcohol	1	10,000	1,00	,900	,095	1
	2	12,000	1,00	-	-	2
	3	12,000	1,00	,700	,145	3
	4	20,000	1,00	,600	,155	4
	5	40,000	,00	-	-	4
	6	100,000	1,00	,480	,164	5
	7	150,000	1,00	,360	,161	6
	8	200,000	1,00	,240	,145	7
	9	300,000	,00	-	-	7
	10	300,000	,00	-	-	7

Grafische vergelijking balansdata (Kaplan-Meier)



SPSS tabellen van de log-rank toets H_0 : beide groepen hebben dezelfde survival

Geen schatting voor de mediaan in de controlegroep
(meer dan 50 % "overleeft" de 5 minuten)

groep	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound	Estimate	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
controle	233,857	33,744	167,719	299,996	-	-	-	-
alcohol	131,400	37,257	58,376	204,424	100,000	88,807	-	274,062
Overall	183,244	27,796	128,764	237,724	200,000	-	-	-

^a. Estimation is limited to the largest survival time if it is censored.

Interpretatie van het gemiddelde?

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	3.937	1	,047

Test of equality of survival distributions for the different levels of groep.

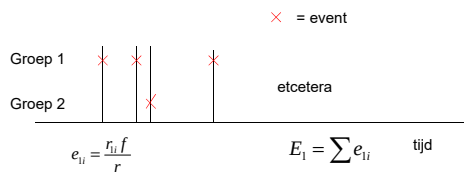
Hoe wordt de Chi-kwadraat berekend?

$P < 0,05$: Significant verschil in overleving tussen de groepen

Berekening van de Chi-kwadraat

- Het is niet goed om een "gewone" kruistabelanalyse op de gegevens te doen; op die manier wordt er geen rekening gehouden met gecensureerde data
- Wel kan er gesommeerd worden over geobserveerde en verwachte aantallen voor ieder moment dat een event plaatsvindt

Achtergrond Log-rank toets



$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

Chi-kwadraat met 1 df

$$\text{of: } X^2 = \frac{(O_1 - E_1)^2}{V}$$

Berekening van de log-rank toets

Time	r1	r2	r	f1	f2	f	e1=r1f/r	f1 - e1	$v = \frac{r_1 f (r - f)}{r^2 (r - 1)}$
10	10	10	20	0	1	1	0,5	-0,5	0,25
12	10	9	19	0	2	2	1,05	-1,05	0,47
20	10	7	17	0	1	1	0,59	-0,59	0,24
25	10	6	16	1	0	1	0,63	0,37	0,23
40*	9	6	15						
45	9	5	14	1	0	1	0,64	0,36	0,23
100	8	5	13	0	1	1	0,62	-0,62	0,24
110*	8	4	12						
150	7	4	11	0	1	1	0,64	-0,64	0,23
185	7	3	10	1	0	1	0,7	0,3	0,21
200	6	3	9	0	1	1	0,67	-0,67	0,22
totaal				3=O	7	10	6,03=E	-3,03	2,32 = V

Chi-kwadraat: $(O-E)^2 / V = (-3,03)^2 / 2,32 = 3,94$

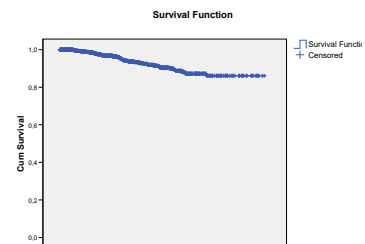
Lifetable, breastcancer data

Gemiddeld over het eerste jaar

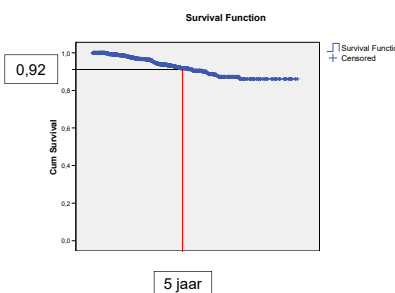
Life Table*

Interval Start Time	Number Entering Interval	Number Withdrawing during Interval	Number Exposed to Risk	Number of Terminal Events	Proportion Terminating	Proportion Surviving	Cumulative Proportion Surviving at End of Interval	Std. Error of Cumulative Proportion Surviving at End of Interval
.000	1207	159	1048	2	.00	1,00	1,00	.00
12,000	1076	183	894	15	.02	.98	.98	.00
24,000	878	147	731	14	.02	.96	.97	.01
36,000	717	166	551	20	.03	.97	.94	.01
48,000	531	153	378	9	.02	.96	.92	.01
60,000	370	121	249	5	.02	.98	.90	.01
72,000	244	91	153	7	.04	.96	.87	.02
84,000	146	59	87	0	.00	1,00	.87	.02
96,000	87	39	48	1	.01	.99	.86	.02
108,000	47	25	22	0	.00	1,00	.86	.02

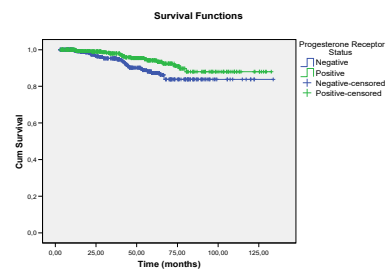
Kaplan-Meier curve (breastcancer data)



5 jaars overleving



Survival curves per subgroep



Vergelijken van de 3-jaars overleving van twee groepen

- Voer een Kaplan-Meier survival uit en haal uit de overlevingstabel de kans (per groep) om 36 maanden te overleven en de bijbehorende standaarddeviaties.
- In dit geval:

Groep	P(T > 36)	SD
≤ 2 cm	0,977	0,006
2 – 5 cm	0,923	0,019

Het 95 % betrouwbaarheidsinterval voor het verschil:

$$p_1 - p_2 \pm 1,96 \cdot \sqrt{s_1^2 + s_2^2} = 0,977 - 0,923 \pm 1,96 \cdot \sqrt{0,006^2 + 0,019^2} = 0,054 \pm 0,039 = [0,015 ; 0,093]$$

Vergelijken van twee groepen met betrekking tot de survival

- “normale toetsen” niet geschikt vanwege censurering

Log rank toets

Case Processing Summary				
Progesterone Receptor Status	Total N	N of Events	Censored	
Negative	389	31	358	92,0%
Positive	462	20	442	95,7%
Overall	851	51	800	94,0%

Means and Medians for Survival Time							
Progesterone Receptor Status	Mean ^a			Median			
	Estimate	Std. Error	95% Confidence Interval	Estimate	Std. Error	95% Confidence Interval	
Negative	119,192	2,527	114,238 124,145	-	-	- -	
Positive	123,485	2,096	119,379 127,599	-	-	- -	
Overall	122,055	1,838	118,845 125,265	-	-	- -	

a. Estimation is limited to the largest survival time if it is censored.

Overall Comparisons

Log Rank (Mantel-Cox)	Chi-Square	df	Sig.
	5,143	1	,023

Test of equality of survival distributions for the different levels of Progesterone Receptor Status.

Meerdere verklarende variabelen: Cox-regressie

- Ook genaamd: proportional hazards model
- Hazard: kans op een “event” in een kleine tijdsperiode

$$h(t) = h_0(t) * e^{b_1 X_1 + b_2 X_2 + \dots + b_k X_k}$$

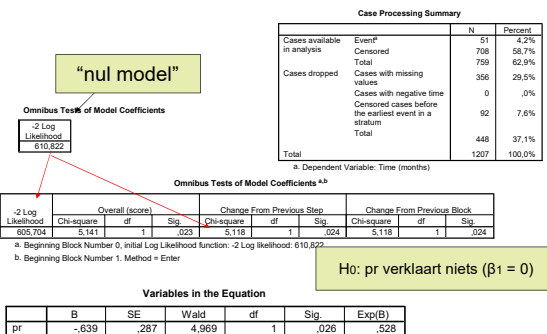
$h_0(t)$: “baseline hazard”

- Cumulatieve hazard

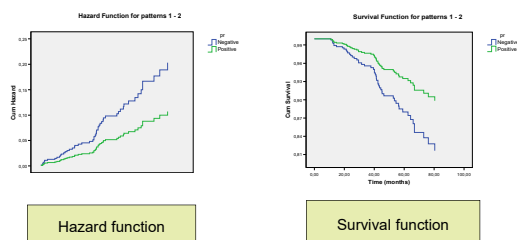
$$H(t) = H_0(t) * e^{b_1 X_1 + b_2 X_2 + \dots + b_k X_k}$$
- Relatie met survival

$$S(t) = e^{-H(t)}$$

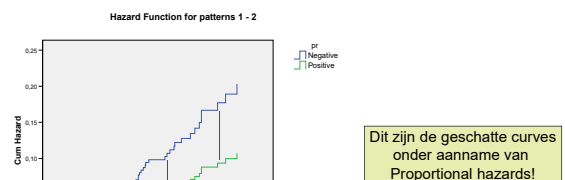
Cox regressie, verklarende variabele progesteron receptor (0 = nee, 1 = ja)



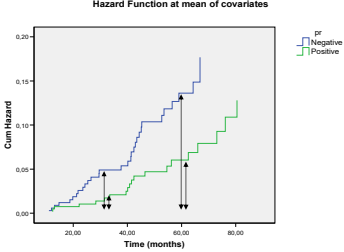
Twee grafieken



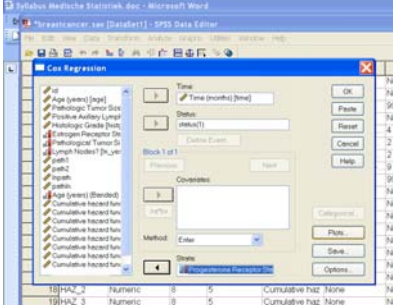
Proportional hazards



Controle van de voorwaarde



In SPSS



Meerdere verklarende variabelen

Omnibus Tests of Model Coefficients

-2 Log Likelihood
610,822

Omnibus Tests of Model Coefficients ^{a,b}

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
598.815	11,598	1	,001	12,008	1	,001	12,008	1	,001

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 610,822

b. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
age	-.039	.012	11.235	1	.001	.962

Omnibus Tests of Model Coefficients ^{a,b}

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
595,895	14,671	2	,001	2,920	1	,088	2,920	1	,088

a. Beginning Block Number 0, initial Log Likelihood function: -2 Log likelihood: 610,822

b. Beginning Block Number 2. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
age	-,035	,012	9,230	1	,002	,965
pr	-,491	,291	2,850	1	,091	,612