

Μοντέλα Αξιοπιστίας και Επιβίωσης

Σειρά 3

Κοντοθανάση Σωτήρια

AM: 09120080

(I)

Γράψτε τη λογαριθμοποιημένη συνάρτηση πιθανοφάνειας για το μοντέλο παλινδρόμησης της Λογαριθμο-κανονικής κατανομής, λαμβάνοντας υπόψη και δεξιά αποκομμένες παρατηρήσεις.

Εάν

$$T_x \sim \text{Lognormal} \rightarrow \varepsilon \sim \text{Normal}$$

$$\ln(T_x) = \mu_0 + \beta'x + \sigma_\varepsilon \sim N(\mu_0 + \beta'x, \sigma_\varepsilon^2)$$

$$\varepsilon \sim N(0,1)$$

$$S_{T_x}(t) = S_\varepsilon(\varepsilon) = S_\varepsilon\left[\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right] = 1 - \Phi\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right)$$

και

$$f_{T_x}(t) = -\frac{dS_{T_x}(t)}{dt} = \frac{1}{\sigma t} f_\varepsilon(\varepsilon) = \frac{1}{\sigma t} f_\varepsilon\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right) = \frac{1}{\sigma t} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right\} \\ \sim \text{Lognormal}(\mu_0 + \beta'x, \sigma^2)$$

Λογαριθμοποιημένη συνάρτηση πιθανοφάνειας για το μοντέλο παλινδρόμησης:

$$L = \prod_{i=1}^n f(t_i, x_i)^{\delta_i} (S(t_i, x_i))^{1-\delta_i}$$

$$L = \prod_{i=1}^n \left(\frac{1}{\sigma t}\right)^{\delta_i} (f_\varepsilon(\varepsilon))^{\delta_i} (S_\varepsilon(\varepsilon))^{1-\delta_i}$$

$$L = \prod_{i=1}^n \left(\frac{1}{\sigma t}\right)^{\delta_i} \left(\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right\}\right)^{\delta_i} \left(1 - \Phi\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right)\right)^{1-\delta_i}$$

$$L = \prod_{i=1}^n \left(\frac{1}{\sigma t}\right)^{\delta_i} \left(\frac{1}{\sqrt{2\pi}}\right)^{\delta_i} \left(\exp\left\{-\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right\}\right)^{\delta_i} \left(1 - \Phi\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right)\right)^{1-\delta_i}$$

$$\ln(L) = \ln\left(\prod_{i=1}^n \left(\frac{1}{\sigma t}\right)^{\delta_i} \left(\frac{1}{\sqrt{2\pi}}\right)^{\delta_i} \left(\exp\left\{-\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right\}\right)^{\delta_i} \left(1 - \Phi\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right)\right)^{1-\delta_i}\right)$$

$$\ln(L) = \sum_{i=0}^n \ln\left(\left(\frac{1}{\sigma t}\right)^{\delta_i} \left(\frac{1}{\sqrt{2\pi}}\right)^{\delta_i} \left(\exp\left\{-\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right\}\right)^{\delta_i} \left(1 - \Phi\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right)\right)^{1-\delta_i}\right)$$

$$\ln(L) = \sum_{i=0}^n \ln\left(\frac{1}{\sigma t}\right)^{\delta_i} + \ln\left(\frac{1}{\sqrt{2\pi}}\right)^{\delta_i} \\ + \ln\left(\exp\left\{-\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right\}\right)^{\delta_i} + \ln\left(1 - \Phi\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right)\right)^{1-\delta_i}$$

$$\ln(L) = \sum_{i=0}^n \delta_i \ln\left(\frac{1}{\sigma t}\right) + \delta_i \ln\left(\frac{1}{\sqrt{2\pi}}\right) \\ + \delta_i \left(-\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right) + (1 - \delta_i) \ln\left(1 - \Phi\left(\frac{\ln t - \mu_0 - \beta'x}{\sigma}\right)\right)$$

(II) Το αρχείο «lung-cancer.txt» περιλαμβάνει τα ακόλουθα δεδομένα:

time = επιβίωση σε ημέρες

status = αποκομμένη (=0) ή μη (=1) παρατήρηση.

και συμμεταβλητές

age: Ηλικία σε έτη

treatment: Τύπος θεραπείας, τύπος 1=0, τύπος 2=1

cell.type:

1=squamous (πλακώδη κύτταρα),

2=small cell (μικρά κύτταρα),

3=adeno (αδενοκαρκίνωμα)

4=large (μεγάλα κύτταρα)

karno.score: Αξιολόγηση επιδόσεων του ασθενή (κακή=0 έως καλή=100)

months: Μήνες από τη διάγνωση της ασθένειας

prior.therapy: Θεραπεία που προηγήθηκε, όχι=0, ναι=1

```
> library(survival)
```

```
> cdat<-read.table("C:\\Users\\Admin\\Desktop\\8ο ΕΞΑΜΗΜΟ\\ΜΟΝΤΕΛΑ ΑΞΙΟΠΙΣΤΙΑΣ ΚΑΙ  
ΕΠΙΒΙΩΣΣ\\σειρα 3\\lung-cancer.txt",header=TRUE)
```

```
> attach(cdat)
```

```
> cdat
```

	treatment	cell.type	time	status	karno.score	months	age	prior.therapy	id
1	1	1	72	1	60	7	69	0	1
2	1	1	411	1	70	5	64	1	2
3	1	1	228	1	60	3	38	0	3
4	1	1	126	1	60	9	63	1	4
5	1	1	118	1	70	11	65	1	5
6	1	1	10	1	20	5	49	0	6
7	1	1	82	1	40	10	69	1	7
8	1	1	110	1	80	29	68	0	8
9	1	1	314	1	50	18	43	0	9
10	1	1	100	0	70	6	70	0	10
11	1	1	42	1	60	4	81	0	11
12	1	1	8	1	40	58	63	1	12
13	1	1	144	1	30	4	63	0	13
14	1	1	25	0	80	9	52	1	14
15	1	1	11	1	70	11	48	1	15
16	1	2	30	1	60	3	61	0	16
17	1	2	384	1	60	9	42	0	17
18	1	2	4	1	40	2	35	0	18
19	1	2	54	1	80	4	63	1	19
20	1	2	13	1	60	4	56	0	20
21	1	2	123	0	40	3	55	0	21
22	1	2	97	0	60	5	67	0	22
23	1	2	153	1	60	14	63	1	23
24	1	2	59	1	30	2	65	0	24
25	1	2	117	1	80	3	46	0	25
26	1	2	16	1	30	4	53	1	26
27	1	2	151	1	50	12	69	0	27
28	1	2	22	1	60	4	68	0	28
29	1	2	56	1	80	12	43	1	29
30	1	2	21	1	40	2	55	1	30
31	1	2	18	1	20	15	42	0	31

32	1	2	139	1	80	2	64	0	32
33	1	2	20	1	30	5	65	0	33
34	1	2	31	1	75	3	65	0	34
35	1	2	52	1	70	2	55	0	35
36	1	2	287	1	60	25	66	1	36
37	1	2	18	1	30	4	60	0	37
38	1	2	51	1	60	1	67	0	38
39	1	2	122	1	80	28	53	0	39
40	1	2	27	1	60	8	62	0	40
41	1	2	54	1	70	1	67	0	41
42	1	2	7	1	50	7	72	0	42
43	1	2	63	1	50	11	48	0	43
44	1	2	392	1	40	4	68	0	44
45	1	2	10	1	40	23	67	1	45
46	1	3	8	1	20	19	61	1	46
47	1	3	92	1	70	10	60	0	47
48	1	3	35	1	40	6	62	0	48
49	1	3	117	1	80	2	38	0	49
50	1	3	132	1	80	5	50	0	50
51	1	3	12	1	50	4	63	1	51
52	1	3	162	1	80	5	64	0	52
53	1	3	3	1	30	3	43	0	53
54	1	3	95	1	80	4	34	0	54
55	1	4	177	1	50	16	66	1	55
56	1	4	162	1	80	5	62	0	56
57	1	4	216	1	50	15	52	0	57
58	1	4	553	1	70	2	47	0	58
59	1	4	278	1	60	12	63	0	59
60	1	4	12	1	40	12	68	1	60
61	1	4	260	1	80	5	45	0	61
62	1	4	200	1	80	12	41	1	62
63	1	4	156	1	70	2	66	0	63
64	1	4	182	0	90	2	62	0	64
65	1	4	143	1	90	8	60	0	65
66	1	4	105	1	80	11	66	0	66
67	1	4	103	1	80	5	38	0	67
68	1	4	250	1	70	8	53	1	68
69	1	4	100	1	60	13	37	1	69
70	2	1	999	1	90	12	54	1	70
71	2	1	112	1	80	6	60	0	71
72	2	1	87	0	80	3	48	0	72
73	2	1	231	0	50	8	52	1	73
74	2	1	242	1	50	1	70	0	74
75	2	1	991	1	70	7	50	1	75
76	2	1	111	1	70	3	62	0	76
77	2	1	1	1	20	21	65	1	77
78	2	1	587	1	60	3	58	0	78
79	2	1	389	1	90	2	62	0	79
80	2	1	33	1	30	6	64	0	80
81	2	1	25	1	20	36	63	0	81
82	2	1	357	1	70	13	58	0	82
83	2	1	467	1	90	2	64	0	83
84	2	1	201	1	80	28	52	1	84
85	2	1	1	1	50	7	35	0	85
86	2	1	30	1	70	11	63	0	86
87	2	1	44	1	60	13	70	1	87
88	2	1	283	1	90	2	51	0	88


```
> #Kaplan Meier
> outp<-survfit(Surv(time,status)~cell.type)
> summary(outp)
```

Call: survfit(formula = Surv(time, status) ~ cell.type)

cell.type=1

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	35	2	0.943	0.0392	0.8690	1.000
8	33	1	0.914	0.0473	0.8261	1.000
10	32	1	0.886	0.0538	0.7863	0.998
11	31	1	0.857	0.0591	0.7487	0.981
15	30	1	0.829	0.0637	0.7127	0.963
25	29	1	0.800	0.0676	0.6779	0.944
30	27	1	0.770	0.0713	0.6426	0.924
33	26	1	0.741	0.0745	0.6083	0.902
42	25	1	0.711	0.0772	0.5749	0.880
44	24	1	0.681	0.0794	0.5423	0.856
72	23	1	0.652	0.0813	0.5105	0.832
82	22	1	0.622	0.0828	0.4793	0.808
110	19	1	0.589	0.0847	0.4448	0.781
111	18	1	0.557	0.0861	0.4112	0.754
112	17	1	0.524	0.0870	0.3784	0.726
118	16	1	0.491	0.0875	0.3464	0.697
126	15	1	0.458	0.0876	0.3152	0.667
144	14	1	0.426	0.0873	0.2849	0.636
201	13	1	0.393	0.0865	0.2553	0.605
228	12	1	0.360	0.0852	0.2265	0.573
242	10	1	0.324	0.0840	0.1951	0.539
283	9	1	0.288	0.0820	0.1650	0.503
314	8	1	0.252	0.0793	0.1362	0.467
357	7	1	0.216	0.0757	0.1088	0.429
389	6	1	0.180	0.0711	0.0831	0.391
411	5	1	0.144	0.0654	0.0592	0.351
467	4	1	0.108	0.0581	0.0377	0.310
587	3	1	0.072	0.0487	0.0192	0.271
991	2	1	0.036	0.0352	0.0053	0.245
999	1	1	0.000	NaN	NA	NA

cell.type=2

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
2	48	1	0.9792	0.0206	0.93958	1.000
4	47	1	0.9583	0.0288	0.90344	1.000
7	46	2	0.9167	0.0399	0.84172	0.998
8	44	1	0.8958	0.0441	0.81345	0.987
10	43	1	0.8750	0.0477	0.78627	0.974
13	42	2	0.8333	0.0538	0.73430	0.946
16	40	1	0.8125	0.0563	0.70926	0.931
18	39	2	0.7708	0.0607	0.66065	0.899
20	37	2	0.7292	0.0641	0.61369	0.866
21	35	2	0.6875	0.0669	0.56812	0.832
22	33	1	0.6667	0.0680	0.54580	0.814
24	32	1	0.6458	0.0690	0.52377	0.796
25	31	2	0.6042	0.0706	0.48052	0.760
27	29	1	0.5833	0.0712	0.45928	0.741
29	28	1	0.5625	0.0716	0.43830	0.722
30	27	1	0.5417	0.0719	0.41756	0.703
31	26	1	0.5208	0.0721	0.39706	0.683

51	25	2	0.4792	0.0721	0.35678	0.644
52	23	1	0.4583	0.0719	0.33699	0.623
54	22	2	0.4167	0.0712	0.29814	0.582
56	20	1	0.3958	0.0706	0.27908	0.561
59	19	1	0.3750	0.0699	0.26027	0.540
61	18	1	0.3542	0.0690	0.24171	0.519
63	17	1	0.3333	0.0680	0.22342	0.497
80	16	1	0.3125	0.0669	0.20541	0.475
87	15	1	0.2917	0.0656	0.18768	0.453
95	14	1	0.2708	0.0641	0.17026	0.431
99	12	2	0.2257	0.0609	0.13302	0.383
117	9	1	0.2006	0.0591	0.11267	0.357
122	8	1	0.1755	0.0567	0.09316	0.331
139	6	1	0.1463	0.0543	0.07066	0.303
151	5	1	0.1170	0.0507	0.05005	0.274
153	4	1	0.0878	0.0457	0.03163	0.244
287	3	1	0.0585	0.0387	0.01600	0.214
384	2	1	0.0293	0.0283	0.00438	0.195
392	1	1	0.0000	NaN	NA	NA

cell.type=3

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
3	27	1	0.9630	0.0363	0.89430	1.000
7	26	1	0.9259	0.0504	0.83223	1.000
8	25	2	0.8519	0.0684	0.72786	0.997
12	23	1	0.8148	0.0748	0.68071	0.975
18	22	1	0.7778	0.0800	0.63576	0.952
19	21	1	0.7407	0.0843	0.59259	0.926
24	20	1	0.7037	0.0879	0.55093	0.899
31	19	1	0.6667	0.0907	0.51059	0.870
35	18	1	0.6296	0.0929	0.47146	0.841
36	17	1	0.5926	0.0946	0.43344	0.810
45	16	1	0.5556	0.0956	0.39647	0.778
48	15	1	0.5185	0.0962	0.36050	0.746
51	14	1	0.4815	0.0962	0.32552	0.712
52	13	1	0.4444	0.0956	0.29152	0.678
73	12	1	0.4074	0.0946	0.25850	0.642
80	11	1	0.3704	0.0929	0.22649	0.606
84	9	1	0.3292	0.0913	0.19121	0.567
90	8	1	0.2881	0.0887	0.15759	0.527
92	7	1	0.2469	0.0850	0.12575	0.485
95	6	1	0.2058	0.0802	0.09587	0.442
117	5	1	0.1646	0.0740	0.06824	0.397
132	4	1	0.1235	0.0659	0.04335	0.352
140	3	1	0.0823	0.0553	0.02204	0.307
162	2	1	0.0412	0.0401	0.00608	0.279
186	1	1	0.0000	NaN	NA	NA

cell.type=4

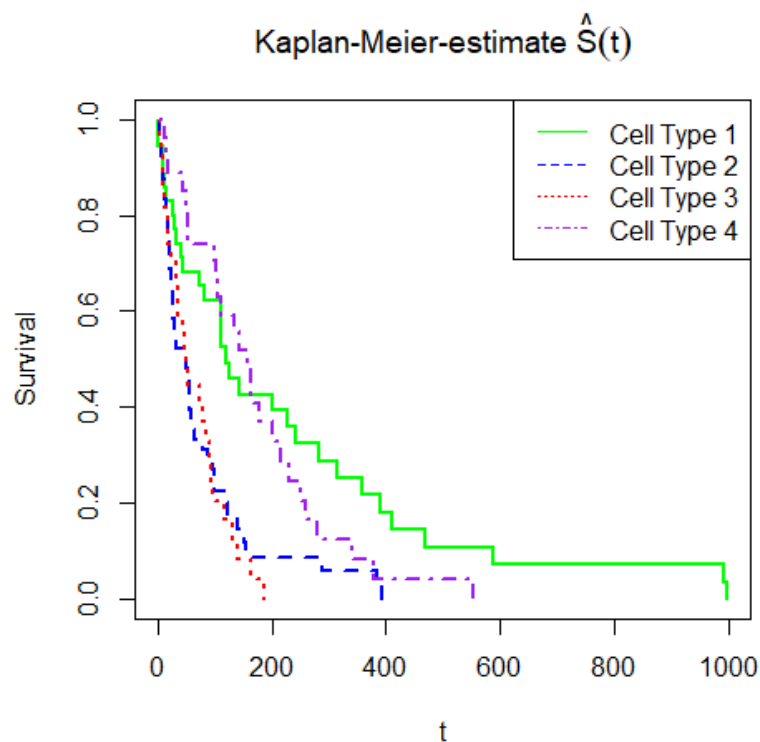
time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
12	27	1	0.9630	0.0363	0.89430	1.000
15	26	1	0.9259	0.0504	0.83223	1.000
19	25	1	0.8889	0.0605	0.77791	1.000
43	24	1	0.8519	0.0684	0.72786	0.997
49	23	1	0.8148	0.0748	0.68071	0.975
52	22	1	0.7778	0.0800	0.63576	0.952
53	21	1	0.7407	0.0843	0.59259	0.926

100	20	1	0.7037	0.0879	0.55093	0.899
103	19	1	0.6667	0.0907	0.51059	0.870
105	18	1	0.6296	0.0929	0.47146	0.841
111	17	1	0.5926	0.0946	0.43344	0.810
133	16	1	0.5556	0.0956	0.39647	0.778
143	15	1	0.5185	0.0962	0.36050	0.746
156	14	1	0.4815	0.0962	0.32552	0.712
162	13	1	0.4444	0.0956	0.29152	0.678
164	12	1	0.4074	0.0946	0.25850	0.642
177	11	1	0.3704	0.0929	0.22649	0.606
200	9	1	0.3292	0.0913	0.19121	0.567
216	8	1	0.2881	0.0887	0.15759	0.527
231	7	1	0.2469	0.0850	0.12575	0.485
250	6	1	0.2058	0.0802	0.09587	0.442
260	5	1	0.1646	0.0740	0.06824	0.397
278	4	1	0.1235	0.0659	0.04335	0.352
340	3	1	0.0823	0.0553	0.02204	0.307
378	2	1	0.0412	0.0401	0.00608	0.279
553	1	1	0.0000	NaN	NA	NA

```

> #Kaplan-Meier GRAPH
> # Sample colors to use for the lines
> colors <- c("green", "blue", "red", "purple")
> # Plot the Kaplan-Meier estimates with different line types and colors
> plot(outp, lty = 1:4, col = colors, main = expression(paste("Kaplan-Meier-
estimate ", hat(S)(t))),
+       xlab = "t", ylab = "Survival", lwd = 2)
> # Add a legend with different line types and colors
> legend("topright", c("Cell Type 1", "Cell Type 2", "Cell Type 3", "Cell Type
4"), lty = 1:4, col = colors)

```



Παρατηρώ ότι η ομάδα με cell type 1 έχει το μεγαλύτερο χρόνο Επιβίωσης από όλες τις υπόλοιπες ομάδες. Η ομάδα με τον μικρότερο χρόνο επιβίωσης είναι η ομάδα με Cell type 3, καθώς τα σκαλακια της K-M πέφτουν πάρα πολύ γρήγορα προς τα κάτω. Ακολουθεί η ομάδα με Cell type 2 με τα αμέσως επόμενα χειρότερα αποτέλεσμα όσον αφορά την επιβίωση και ωστερα η ομάδα με cel type 4.

A) Για τη συμμεταβλητή cell.type, εφαρμόστε τους ελέγχους log-rank και Wilcoxon, καθώς και γραφικές παραστάσεις της Kaplan-Meier. (PART 1)

Log -rank

```
> out1<-survdif(Surv(time,status) ~ cell.type)
> out1
Call:
survdif(formula = Surv(time, status) ~ cell.type)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
cell.type=1	35	31	47.7	5.82	10.53
cell.type=2	48	45	30.1	7.37	10.20
cell.type=3	27	26	15.7	6.77	8.19
cell.type=4	27	26	34.5	2.12	3.02

Chisq= 25.4 on 3 degrees of freedom, **p= 1e-05**

Wilcoxon

```
> out2<-survdif(Surv(time, status) ~ cell.type, rho=1)
> out2
Call:
survdif(formula = Surv(time, status) ~ cell.type, rho = 1)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
cell.type=1	35	13.39	20.2	2.27	5.05
cell.type=2	48	28.43	19.0	4.67	9.33
cell.type=3	27	16.07	10.8	2.56	4.23
cell.type=4	27	9.56	17.5	3.59	7.57

Chisq= 19.7 on 3 degrees of freedom, **p= 2e-04**
>

Το Log-rank τεστ δείχνει ότι υπάρχει σημαντική διαφορά στις κατανομές επιβίωσης μεταξύ των τεσσάρων τύπων κυττάρων ($p=1e-05$). Η πολύ χαμηλή τιμή p σημαίνει ότι απορρίπτουμε την μηδενική υπόθεση (ότι δεν υπάρχει διαφορά μεταξύ των ομάδων) με υψηλή στατιστική σημαντικότητα.

Το Wilcoxon τεστ, επίσης, δείχνει ότι υπάρχει σημαντική διαφορά στις κατανομές επιβίωσης μεταξύ των τεσσάρων τύπων κυττάρων ($p=2e-04$). Και σε αυτή την περίπτωση, η χαμηλή τιμή p σημαίνει ότι απορρίπτουμε τη μηδενική υπόθεση με υψηλή στατιστική σημαντικότητα.

Με βάση αυτή τη συµµεταβλητή εξετάστε γραφικά αν στα δεδοµένα ταιριάζει ένα µοντέλο παλινδρόµησης της επιταχυνόµενης διάρκειας ζωής (AL) ή ένα µοντέλο της αναλογικής διακινδύνευσης (PH) (µε ταυτόχρονο έλεγχο για την κατανοµή Weibull).(PART 2)

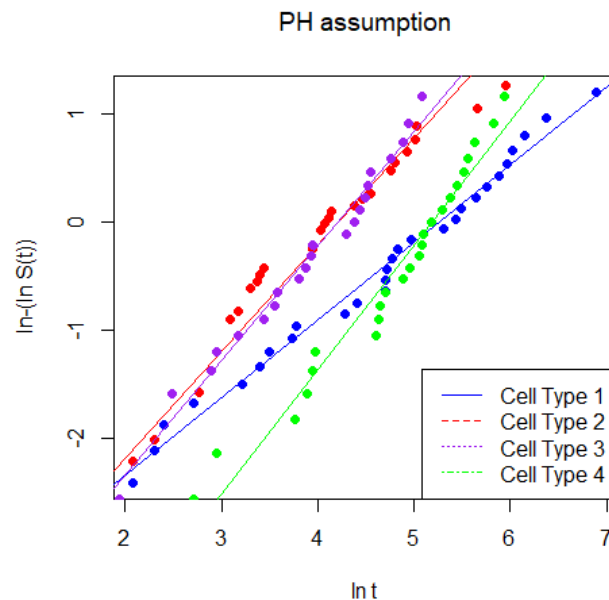
```
> ###Kaplan-Meier for 1st group
> group1<- Surv(time[cell.type=="1"],status[cell.type=="1"])
> outp1<-survfit(group1~1, type="kaplan-meier", data=cdat)
> #Extract values when status=1
> Utime1<- outp1$time[outp1$n.event==1]
> SKM1<-outp1$surv[outp1$n.event==1]
>
> ###Kaplan-Meier for 2nd group
> group2<- Surv(time[cell.type=="2"],status[cell.type=="2"])
> outp2<-survfit(group2~1, type="kaplan-meier", data=cdat)
> #Extract values when status=2
> Utime2<- outp2$time[outp2$n.event==1]
> SKM2<-outp2$surv[outp2$n.event==1]
>
> ###Kaplan-Meier for 3rd group
> group3<- Surv(time[cell.type=="3"],status[cell.type=="3"])
> outp3<-survfit(group3~1, type="kaplan-meier", data=cdat)
> #Extract values when status=3
> Utime3<- outp3$time[outp3$n.event==1]
> SKM3<-outp3$surv[outp3$n.event==1]
>
> ###Kaplan-Meier for 4th group
> group4<- Surv(time[cell.type=="4"],status[cell.type=="4"])
> outp4<-survfit(group4~1, type="kaplan-meier", data=cdat)
> #Extract values when status=4
> Utime4<- outp4$time[outp4$n.event==1]
> SKM4<-outp4$surv[outp4$n.event==1]
```

Ύστερα αφαιρώ τις τελευταίες παρατηρήσεις από τα διανύσµατα Utime1,2,3,4 και SKM1,2,3,4 γιατί η K-M µου δίνει µηδενικές τιµές.

```
># Remove the last element from Utime1 and SKM1
Utime1 <- head(Utime1, -1)
SKM1 <- head(SK1, -1)
> # Remove the last element from Utime1 and SKM1
> Utime1 <- head(Utime1, -1)
> SKM1 <- head(SK1, -1)
>
> # Remove the last element from Utime2 and SKM2
> Utime2 <- head(Utime2, -1)
> SKM2 <- head(SK2, -1)
>
> # Remove the last element from Utime3 and SKM3
> Utime3 <- head(Utime3, -1)
> SKM3 <- head(SK3, -1)
>
> # Remove the last element from Utime4 and SKM4
> Utime4 <- head(Utime4, -1)
> SKM4 <- head(SK4, -1)
```

Κάνω την γραφική παράστασή για το PH μοντέλο .

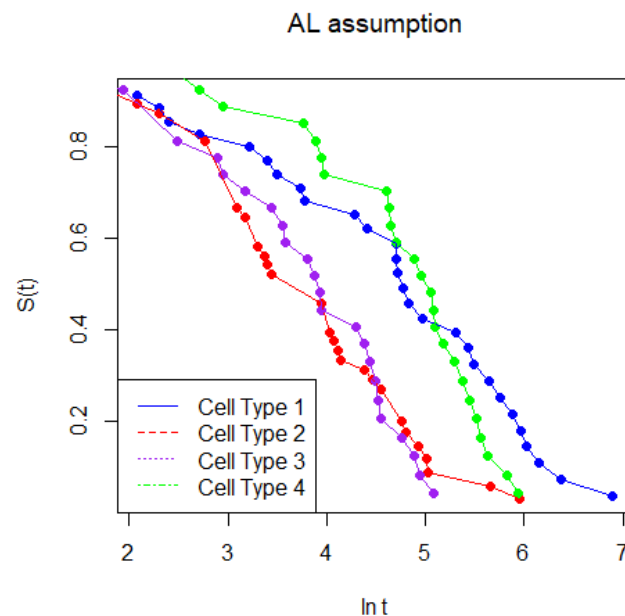
```
> #PH assumption plot
> plot(log(Utime1), log(-log(SKM1)) , main=expression(paste("PH assumption")),
xlab="ln t",
+ ylab="ln-(ln S(t))", col="blue",pch=19)
> abline(lm(log(-log(SKM1))~ log(Utime1)), col="blue")
>
> points(log(Utime2), log(-log(SKM2)), col="red", pch=19)
> abline(lm(log(-log(SKM2))~ log(Utime2)), col="red")
>
> points(log(Utime3), log(-log(SKM3)), col="purple", pch=19)
> abline(lm(log(-log(SKM3))~ log(Utime3)), col="purple")
>
> points(log(Utime4), log(-log(SKM4)), col="green", pch=19)
> abline(lm(log(-log(SKM4))~ log(Utime4)), col="green")
>
> legend("bottomright", c("Cell Type 1", "Cell Type 2", "Cell Type 3", "Cell
Type 4"), col=c("blue", "red", "purple", "green"),
+ lty=1:4)
>
```



Στην γραφική παράσταση του μοντέλου αναλογικής διακινδύνευσης παρατηρούμε ότι οι ευθείες για το Cell type 2 και Cell type 3 σχεδόν ταυτίζονται. Επιπλέον είναι παράλληλες με την ευθεία των δεδομένων για το cell type 4. Θα μπορούσαμε να πούμε ότι το μοντέλο δηλ PH ταιριάζει στα δεδομένα μας, αν δεν υπήρχε η ευθεία του cell type 1 η οποία δεν είναι παράλληλη με καμία από τις υπόλοιπες και τέμνει τις 2 από αυτές

Κάνω την γραφική παράστασή για το AL μοντέλο .

```
> #Examine the AL assumption
> plot(log(Utime1), SKM1, main=expression(paste("AL assumption")), xlab="ln t",
+ ylab="S(t)", col="blue",pch=19,lty = 1:4)
> points(log(Utime2), SKM2,col="red",pch=19)
> points(log(Utime3), SKM3,col="purple",pch=19)
> points(log(Utime4), SKM4,col="green",pch=19)
>
> lines(log(Utime1), SKM1, col="blue")
> lines(log(Utime2), SKM2, col="red")
> lines(log(Utime3), SKM3, col="purple")
> lines(log(Utime4), SKM4, col="green")
> legend("bottomleft", c("Cell Type 1", "Cell Type 2", "Cell Type 3", "Cell Type
4"), col=c("blue", "red", "purple", "green"),
+ lty=1:4)
>
```



Στο παραπάνω γράφημα δεν είναι φανερή κάποια μετατόπιση της γραφικής μου παραστάσης καθώς αλλάζει το Cell type, αντιθέτως βλέπω οι τέσσερις γραφικές παραστάσεις να τέμνονται μεταξύ τους ανά ζευγάρια των 2, το Cell type 1 με το Cell type 4 και το Cell type 3 με το Cell type 2.

Δεν μπορώ να βγάλω συμπέρασμα ποιο μοντέλο ταίριαζει καλύτερα στα δεδομένα μου ανάμεσα στο μοντέλο αναλογικής διακινδυνεύσης PH και το μοντέλο επιταχυνόμενης διάρκειας ζωής AL.

B)

- i. Να προσαρμοστεί ένα μοντέλο παλινδρόμησης της κατανομής Weibull (Χρησιμοποιήστε ελέγχους Wald, ελέγχους του λόγου των πιθανοφανειών, το κριτήριο AIC, καθώς και την Backward τεχνική με βήματα).

```
> mod <- survreg(Surv(time,status)~ treatment + factor(cell.type)+ karno.score +  
months + age + prior.therapy + id, data=cdat, dist="weibull")
```

```
> mod
```

Call:

```
survreg(formula = Surv(time, status) ~ treatment + factor(cell.type) +  
karno.score + months + age + prior.therapy + id, data = cdat,  
dist = "weibull")
```

Coefficients:

(Intercept)	treatment	factor(cell.type) 2	factor(cell.type) 3
2.7297135997	0.4988076251	-0.5935622077	-0.7025378039
factor(cell.type) 4	karno.score	months	age
0.1754697094	0.0301815145	0.0006327761	0.0080822852
prior.therapy	id		
-0.0714614633	-0.0106056098		

Scale= 0.9313627

Loglik(model)= -715.2 Loglik(intercept only)= -748.1
Chisq= 65.7 on 9 degrees of freedom, p= 1.05e-10
n= 137

```
> summary(mod)
```

Call:

```
survreg(formula = Surv(time, status) ~ treatment + factor(cell.type) +  
karno.score + months + age + prior.therapy + id, data = cdat,  
dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	2.729714	1.183843	2.31	0.021
treatment	0.498808	0.942628	0.53	0.597
factor(cell.type) 2	-0.593562	0.388777	-1.53	0.127
factor(cell.type) 3	-0.702538	0.605776	-1.16	0.246
factor(cell.type) 4	0.175470	0.773491	0.23	0.821
karno.score	0.030182	0.004833	6.25	4.2e-10
months	0.000633	0.008473	0.07	0.940
age	0.008082	0.008929	0.91	0.365
prior.therapy	-0.071461	0.216062	-0.33	0.741
id	-0.010606	0.013481	-0.79	0.431
Log(scale)	-0.071107	0.066322	-1.07	0.284

Scale= 0.931

Weibull distribution

Loglik(model)= -715.2 Loglik(intercept only)= -748.1
Chisq= 65.7 on 9 degrees of freedom, p= 1.1e-10
Number of Newton-Raphson Iterations: 6
n= 137

```
> step(mod, direction="backward", test="Chisq")
```

```
Start: AIC=1452.48
```

```
Surv(time, status) ~ treatment + factor(cell.type) + karno.score +  
  months + age + prior.therapy + id
```

	Df	AIC	LRT	Pr(>Chi)
- months	1	1450.5	0.006	0.940217
- prior.therapy	1	1450.6	0.109	0.741457
- treatment	1	1450.8	0.281	0.595799
- id	1	1451.1	0.624	0.429731
- age	1	1451.3	0.804	0.369760
<none>		1452.5		
- factor(cell.type)	3	1461.9	15.420	0.001491 **
- karno.score	1	1485.7	35.216	2.951e-09 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=1450.48
```

```
Surv(time, status) ~ treatment + factor(cell.type) + karno.score +  
  age + prior.therapy + id
```

	Df	AIC	LRT	Pr(>Chi)
- prior.therapy	1	1448.6	0.111	0.73940
- treatment	1	1448.8	0.276	0.59950
- id	1	1449.1	0.621	0.43066
- age	1	1449.3	0.802	0.37056
<none>		1450.5		
- factor(cell.type)	3	1460.0	15.464	0.00146 **
- karno.score	1	1484.5	36.025	1.948e-09 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=1448.6
```

```
Surv(time, status) ~ treatment + factor(cell.type) + karno.score +  
  age + id
```

	Df	AIC	LRT	Pr(>Chi)
- treatment	1	1446.8	0.238	0.625736
- id	1	1447.2	0.577	0.447524
- age	1	1447.4	0.831	0.361980
<none>		1448.6		
- factor(cell.type)	3	1458.0	15.443	0.001475 **
- karno.score	1	1482.5	35.945	2.029e-09 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=1446.83
```

```
Surv(time, status) ~ factor(cell.type) + karno.score + age +  
  id
```

	Df	AIC	LRT	Pr(>Chi)
- age	1	1445.5	0.680	0.4095907
- id	1	1446.8	1.953	0.1622731
<none>		1446.8		
- factor(cell.type)	3	1458.3	17.500	0.0005575 ***
- karno.score	1	1481.0	36.115	1.86e-09 ***

```
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Step: **AIC=1445.51**

```
Surv(time, status) ~ factor(cell.type) + karno.score + id
```

	Df	AIC	LRT	Pr(>Chi)
- id	1	1445.0	1.517	0.2181465
<none>		1445.5		
- factor(cell.type)	3	1456.4	16.913	0.0007365 ***
- karno.score	1	1479.3	35.765	2.226e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Step: **AIC=1445.03**

```
Surv(time, status) ~ factor(cell.type) + karno.score
```

	Df	AIC	LRT	Pr(>Chi)
<none>		1445.0		
- factor(cell.type)	3	1458.1	19.042	0.0002679 ***
- karno.score	1	1478.3	35.285	2.848e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Call:

```
survreg(formula = Surv(time, status) ~ factor(cell.type) + karno.score,  
data = cdat, dist = "weibull")
```

Coefficients:

	(Intercept)	factor(cell.type)2	factor(cell.type)3	factor(cell.type)4
karno.score	3.48063259 0.02917581	-0.70815457	-1.10848938	-0.32200154

Scale= 0.9378158

Loglik(model)= -716.5 Loglik(intercept only)= -748.1
Chisq= 63.15 on 4 degrees of freedom, p= 6.3e-13
n= 137

>
>

Karno.score: Ο παράγοντας karno.score είναι πολύ σημαντικός ($p < 0.001$), με θετικό συντελεστή, υποδεικνύοντας ότι υψηλότερο σκορ Karnofsky σχετίζεται με καλύτερη επιβίωση.

Factor(cell.type):

Οι διαφορετικοί τύποι κυττάρων (cell.type) είναι επίσης σημαντικοί ($p < 0.001$).

Συγκεκριμένα, οι τύποι κυττάρων 2, 3 και 4 έχουν αρνητικούς συντελεστές, υποδεικνύοντας χαμηλότερη επιβίωση σε σχέση με τον τύπο 1 (αναφορά).

Τιμές AIC: Το τελικό μοντέλο έχει χαμηλότερη τιμή AIC (1445.03), υποδεικνύοντας καλύτερη προσαρμογή σε σχέση με το αρχικό μοντέλο.

Σημαντικότητα Μοντέλου: Το συνολικό μοντέλο είναι σημαντικό με πολύ χαμηλή τιμή p ($6.3e-13$), υποδεικνύοντας ότι οι επιλεγμένοι παράγοντες έχουν σημαντική επίδραση στην επιβίωση.


```
> #FINAL MODEL
>
> modfinal <- survreg(Surv(time,status)~ factor(cell.type)+ karno.score,
data=cdat, dist="weibull")
> modfinal
Call:
survreg(formula = Surv(time, status) ~ factor(cell.type) + karno.score,
data = cdat, dist = "weibull")

Coefficients:
      (Intercept)  factor(cell.type)2  factor(cell.type)3  factor(cell.type)4
karno.score
      3.48063259      -0.70815457      -1.10848938      -0.32200154
0.02917581

Scale= 0.9378158

Loglik(model)= -716.5  Loglik(intercept only)= -748.1
      Chisq= 63.15 on 4 degrees of freedom, p= 6.3e-13
n= 137
```

```
> summary(modfinal)
```

```
Call:
survreg(formula = Surv(time, status) ~ factor(cell.type) + karno.score,
data = cdat, dist = "weibull")

      Value Std. Error      z      p
(Intercept)    3.48063    0.34048 10.22 < 2e-16
factor(cell.type)2 -0.70815    0.22609 -3.13  0.0017
factor(cell.type)3 -1.10849    0.25269 -4.39 1.2e-05
factor(cell.type)4 -0.32200    0.25002 -1.29 0.1978
karno.score      0.02918    0.00462  6.31 2.8e-10
Log(scale)      -0.06420    0.06596 -0.97  0.3304

Scale= 0.938
```

```
Weibull distribution
Loglik(model)= -716.5  Loglik(intercept only)= -748.1
      Chisq= 63.15 on 4 degrees of freedom, p= 6.3e-13
Number of Newton-Raphson Iterations: 5
n= 137
```

- ii. Κατασκευάστε 0.95-διαστήματα εμπιστοσύνης για τους συντελεστές των συμμεταβλητών του τελικού μοντέλου και να δώσετε ερμηνείες.

```
> confint.default(modfinal)
```

	2.5 %	97.5 %
(Intercept)	2.81330376	4.14796142
factor(cell.type) 2	-1.15128713	-0.26502201
factor(cell.type) 3	-1.60374377	-0.61323499
factor(cell.type) 4	-0.81202533	0.16802225
karno.score	0.02011407	0.03823756

Ερμηνεία

(Intercept): Το διάστημα δεν περιέχει το 0, επομένως ο συντελεστής είναι στατιστικά σημαντικός.

factor(cell.type)2: Το διάστημα δεν περιέχει το 0, επομένως ο συντελεστής είναι στατιστικά σημαντικός.

factor(cell.type)3: Το διάστημα δεν περιέχει το 0, επομένως ο συντελεστής είναι στατιστικά σημαντικός.

factor(cell.type)4: Το διάστημα περιέχει το 0, επομένως ο συντελεστής δεν είναι στατιστικά σημαντικός.

karno.score: Το διάστημα δεν περιέχει το 0, επομένως ο συντελεστής είναι στατιστικά σημαντικός.

```
> exp(confint.default(modfinal))
```

	2.5 %	97.5 %
(Intercept)	16.6648842	63.3048168
factor(cell.type) 2	0.3162295	0.7671891
factor(cell.type) 3	0.2011421	0.5415960
factor(cell.type) 4	0.4439580	1.1829629
karno.score	1.0203177	1.0389780

Ερμηνεία

(Intercept): Το διάστημα δεν περιέχει το 1, επομένως ο εκθετικός συντελεστής είναι στατιστικά σημαντικός.

factor(cell.type)2: Το διάστημα δεν περιέχει το 1, επομένως ο εκθετικός συντελεστής είναι στατιστικά σημαντικός.

factor(cell.type)3: Το διάστημα δεν περιέχει το 1, επομένως ο εκθετικός συντελεστής είναι στατιστικά σημαντικός.

factor(cell.type)4: Το διάστημα περιέχει το 1, επομένως ο εκθετικός συντελεστής δεν είναι στατιστικά σημαντικός.

karno.score: Το διάστημα δεν περιέχει το 1, επομένως ο εκθετικός συντελεστής είναι στατιστικά σημαντικός.

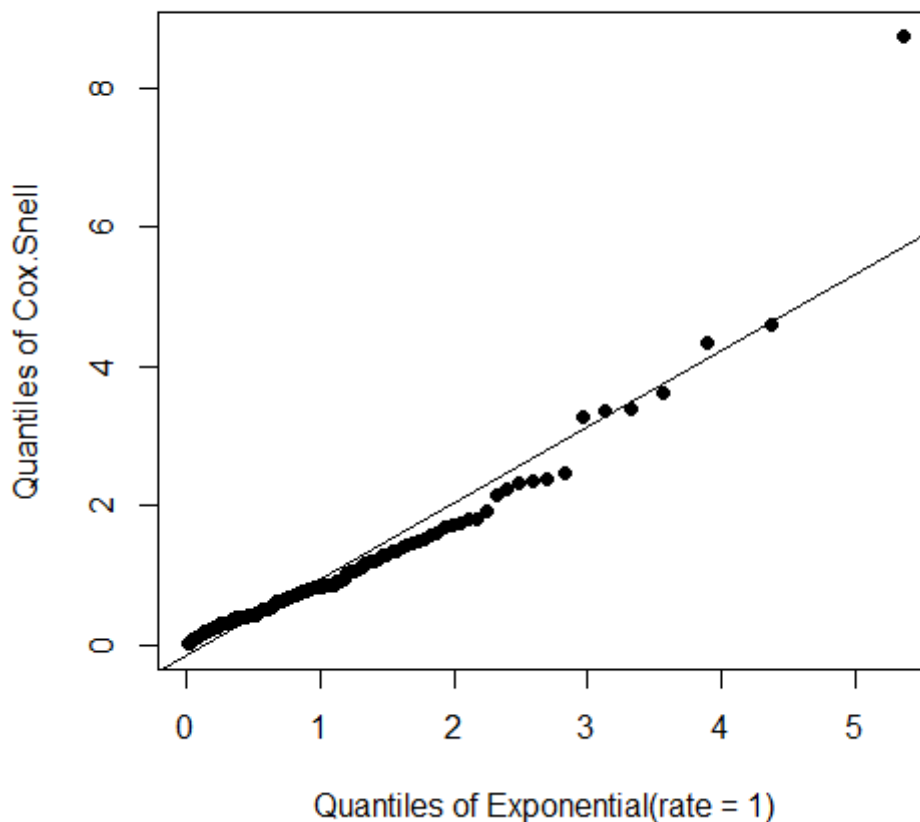
- iii. Επιπλέον να γίνουν οι γραφικές παραστάσεις των υπολοίπων Cox-Snell και να δωθούν ερμηνείες και συμπεράσματα.

```
> standres<-(log(time)-modfinal$linear.predictors)/modfinal$scale
> csres<-exp(standres)
> csres
[1] 0.361386656 1.696418492 1.235287846 0.656334773 0.448370640 0.152839228 0.773428226 0.304800355
[9] 2.371865598 0.375827748 0.203407718 0.064666231 1.924397472 0.062790947 0.035712368 0.302336337
[17] 4.582649845 0.065709376 0.303716182 0.123945572 2.535874737 1.056658111 1.717820392 1.581356468
[25] 0.692668725 0.393295656 2.312038500 0.217200212 0.315725368 0.385069065 0.608659177 0.832369656
[33] 0.498947702 0.196340087 0.398200144 3.359563115 0.445926693 0.532377592 0.724276875 0.270208366
[41] 0.414551640 0.087431905 0.910302871 8.727436550 0.174563633 0.392853784 1.121263050 1.017394206
[49] 1.061493199 1.207199395 0.238048946 1.501818866 0.101136186 0.850073800 1.814438206 0.649234731
[57] 2.243659632 3.281545356 2.151315100 0.140462746 1.075185781 0.812802177 0.851207076 0.538517327
[65] 0.416408503 0.408873173 0.400573981 1.407445021 0.723128896 2.347532633 0.310713187 0.237348811
[73] 1.709749332 1.796699576 4.336203731 0.420065559 0.013119803 3.386134001 0.858687486 0.399962661
[81] 0.406032936 1.459832814 1.043434186 0.579666731 0.005159352 0.104097260 0.213752130 0.611660887
[89] 0.092612617 0.632981333 0.825310965 0.717397799 0.315183250 0.940911582 0.031378833 0.498947702
[97] 0.222332041 0.238316697 0.791179382 0.039643782 0.496145326 0.472089969 0.182367769 0.757139142
[105] 1.174396420 1.353791807 0.543265639 0.680405033 0.500661737 0.407579320 0.257540339 0.815852042
[113] 1.494995795 0.832920890 1.195891184 0.154489450 0.412290036 3.623182270 0.182882555 1.754438898
[121] 1.274920473 0.745535883 0.388572236 1.330058765 2.456501652 0.360070833 0.897831058 0.312951736
[129] 0.367459094 0.243224734 0.294022368 1.431245973 0.614629777 0.808246718 1.293681011 1.602426658
[137] 0.859412326

> Cox.Snell<- csres [status == 1]

> library(EnvStats)
> qqPlot(Cox.Snell, distribution="exp", param.list = list(rate=1),
add.line=TRUE, pch=19)
```

Exponential Q-Q Plot for Cox.Snell



Γ) Στα ίδια δεδομένα

- i. Να προσαρμοστεί το μοντέλο του Cox (χρησιμοποιήστε πάλι ελέγχους Wald, του λόγου των πιθανοφανειών, το κριτήριο AIC και την Backward τεχνική με βήματα).

Ελέγχους Wald

```
> mod2<- coxph(Surv(time,status)~treatment +factor(cell.type)+karno.score +  
months + age + prior.therapy + id,ties=c("breslow"))  
> mod2
```

Call:

```
coxph(formula = Surv(time, status) ~ treatment + factor(cell.type) +  
karno.score + months + age + prior.therapy + id, ties = c("breslow"))
```

	coef	exp(coef)	se(coef)	z	p
treatment	-0.586327	0.556367	1.020826	-0.574	0.566
factor(cell.type)2	0.575239	1.777556	0.426576	1.349	0.177
factor(cell.type)3	0.665193	1.944867	0.669090	0.994	0.320
factor(cell.type)4	-0.289940	0.748309	0.837381	-0.346	0.729
karno.score	-0.032719	0.967810	0.005482	-5.969	2.39e-09
months	-0.001508	0.998493	0.009234	-0.163	0.870
age	-0.010762	0.989295	0.009632	-1.117	0.264
prior.therapy	0.106091	1.111923	0.235727	0.450	0.653
id	0.012759	1.012841	0.014573	0.876	0.381

Likelihood ratio test=62.17 on 9 df, p=5.099e-10
n= 137, number of events= 128

```
> summary(mod2)
```

```
Call:
```

```
coxph(formula = Surv(time, status) ~ treatment + factor(cell.type) +  
      karno.score + months + age + prior.therapy + id, ties = c("breslow"))
```

```
n= 137, number of events= 128
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
treatment	-0.586327	0.556367	1.020826	-0.574	0.566
factor(cell.type)2	0.575239	1.777556	0.426576	1.349	0.177
factor(cell.type)3	0.665193	1.944867	0.669090	0.994	0.320
factor(cell.type)4	-0.289940	0.748309	0.837381	-0.346	0.729
karno.score	-0.032719	0.967810	0.005482	-5.969	2.39e-09 ***
months	-0.001508	0.998493	0.009234	-0.163	0.870
age	-0.010762	0.989295	0.009632	-1.117	0.264
prior.therapy	0.106091	1.111923	0.235727	0.450	0.653
id	0.012759	1.012841	0.014573	0.876	0.381

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
treatment	0.5564	1.7974	0.07524	4.1142
factor(cell.type)2	1.7776	0.5626	0.77041	4.1013
factor(cell.type)3	1.9449	0.5142	0.52403	7.2181
factor(cell.type)4	0.7483	1.3363	0.14498	3.8624
karno.score	0.9678	1.0333	0.95747	0.9783
months	0.9985	1.0015	0.98058	1.0167
age	0.9893	1.0108	0.97079	1.0081
prior.therapy	1.1119	0.8993	0.70052	1.7649
id	1.0128	0.9873	0.98432	1.0422

```
Concordance= 0.741 (se = 0.021 )
```

```
Likelihood ratio test= 62.17 on 9 df, p=5e-10
```

```
Wald test = 62.84 on 9 df, p=4e-10
```

```
Score (logrank) test = 67.44 on 9 df, p=5e-11
```

```
>
```

Backward τεχνική με βήματα

```
> mod3<-step(mod2, direction="backward", test="Chisq")
```

```
Start: AIC=967.59
```

```
Surv(time, status) ~ treatment + factor(cell.type) + karno.score +  
months + age + prior.therapy + id
```

	Df	AIC	LRT	Pr(>Chi)
- months	1	965.62	0.027	0.869176
- prior.therapy	1	965.79	0.201	0.654211
- treatment	1	965.92	0.329	0.566235
- id	1	966.36	0.765	0.381794
- age	1	966.82	1.221	0.269075
<none>		967.59		
- factor(cell.type)	3	976.15	14.560	0.002234 **
- karno.score	1	1000.61	35.013	3.276e-09 ***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=965.62
```

```
Surv(time, status) ~ treatment + factor(cell.type) + karno.score +  
age + prior.therapy + id
```

	Df	AIC	LRT	Pr(>Chi)
- prior.therapy	1	963.80	0.176	0.674794
- treatment	1	963.93	0.308	0.578820
- id	1	964.36	0.738	0.390337
- age	1	964.82	1.194	0.274447
<none>		965.62		
- factor(cell.type)	3	974.17	14.547	0.002247 **
- karno.score	1	999.39	35.768	2.223e-09 ***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=963.8
```

```
Surv(time, status) ~ treatment + factor(cell.type) + karno.score +  
age + id
```

	Df	AIC	LRT	Pr(>Chi)
- treatment	1	962.06	0.261	0.609495
- id	1	962.48	0.679	0.409876
- age	1	963.03	1.230	0.267327
<none>		963.80		
- factor(cell.type)	3	972.20	14.407	0.002401 **
- karno.score	1	997.47	35.671	2.337e-09 ***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Step: AIC=962.06
```

```
Surv(time, status) ~ factor(cell.type) + karno.score + age +  
id
```

	Df	AIC	LRT	Pr(>Chi)
- age	1	961.12	1.059	0.303425
<none>		962.06		
- id	1	962.58	2.521	0.112365
- factor(cell.type)	3	971.75	15.692	0.001311 **

```
- karno.score          1 995.79 35.735 2.26e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step: **AIC=961.12**

```
Surv(time, status) ~ factor(cell.type) + karno.score + id
```

	Df	AIC	LRT	Pr(>Chi)
- id	1	961.00	1.880	0.170317
<none>		961.12		
- factor(cell.type)	3	970.10	14.982	0.001832 **
- karno.score	1	993.91	34.796	3.662e-09 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step: **AIC=961**

```
Surv(time, status) ~ factor(cell.type) + karno.score
```

	Df	AIC	LRT	Pr(>Chi)
<none>		961.00		
- factor(cell.type)	3	972.14	17.145	0.00066 ***
- karno.score	1	993.20	34.200	4.973e-09 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```


- ii. Να κατασκευαστούν 0.95-διαστήματα εμπιστοσύνης για τους συντελεστές των συµµεταβλητών του τελικού µοντέλου και να δώσετε ερµηνείες.

Και άρα το τελικό µοντέλο µε όλες τις µεταβλητές που είναι στατιστικά σηµαντικές είναι το παρακάτω :

```
> #Τελικό μοντέλο
>
> mod4<- coxph(Surv(time,status)~factor(cell.type)+karno.score ,
ties=c("breslow"))
> mod4
Call:
coxph(formula = Surv(time, status) ~ factor(cell.type) + karno.score,
      ties = c("breslow"))

              coef exp(coef)    se(coef)      z      p
factor(cell.type)2  0.712148  2.038365  0.252740  2.818  0.00484
factor(cell.type)3  1.150801  3.160725  0.292861  3.930 8.51e-05
factor(cell.type)4  0.325143  1.384228  0.276694  1.175  0.23996
karno.score         -0.030904  0.969569  0.005179 -5.968 2.41e-09

Likelihood ratio test=58.77 on 4 df, p=5.257e-12
n= 137, number of events= 128
>
```

```
> summary(mod4)
Call:
coxph(formula = Surv(time, status) ~ factor(cell.type) + karno.score,
      ties = c("breslow"))

n= 137, number of events= 128

              coef exp(coef)    se(coef)      z Pr(>|z|)
factor(cell.type)2  0.712148  2.038365  0.252740  2.818  0.00484 **
factor(cell.type)3  1.150801  3.160725  0.292861  3.930 8.51e-05 ***
factor(cell.type)4  0.325143  1.384228  0.276694  1.175  0.23996
karno.score         -0.030904  0.969569  0.005179 -5.968 2.41e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
factor(cell.type)2	2.0384	0.4906	1.2421	3.3451
factor(cell.type)3	3.1607	0.3164	1.7803	5.6114
factor(cell.type)4	1.3842	0.7224	0.8048	2.3808
karno.score	0.9696	1.0314	0.9598	0.9795

```
Concordance= 0.734 (se = 0.023 )
Likelihood ratio test= 58.77 on 4 df, p=5e-12
Wald test              = 60.58 on 4 df, p=2e-12
Score (logrank) test = 63.22 on 4 df, p=6e-13
```

- iii. Να εξεταστεί αν το τελικό μοντέλο πληρεί τις προϋποθέσεις της αναλογικότητας της διακινδύνευσης - μέσω στατιστικών ελέγχων, καθώς και με τη χρήση υπολοίπων Schoenfeld, δηλαδή να εξεταστεί αν οι συντελεστές των μεταβλητών δε σχετίζονται με το χρόνο (αν $\beta_j(t)=\beta_j$) και στο τέλος πάλι ερμηνείες και συμπεράσματα. (έχω συσχετίσή με τον χρόνο η όχι).

```
> sresid<-resid(mod4,type="scaledsch")
```

```
> sresid
```

	factor(cell.type)2	factor(cell.type)3	factor(cell.type)4	karno.score
1	-5.75917314	-5.475703976	-3.943312322	-0.1262046601
1	-4.95950337	-5.396808093	-4.478648296	-0.0232229975
2	2.85938123	-0.116328756	0.263533552	-0.0318569016
3	-0.15571326	5.418001240	0.695423805	-0.0902209659
4	2.86249026	-0.023916027	0.272112979	-0.0333567359
7	3.15490716	-0.011744712	0.093029913	0.0009409872
7	2.35523739	-0.090640595	0.628365887	-0.1020406754
7	0.11492135	5.537274206	0.525634356	-0.0574120299
8	-5.23843946	-5.394832398	-4.347712337	-0.0619366555
8	-0.36863223	5.530808307	0.891423240	-0.1279612345
8	4.00413653	0.113282527	-0.433407826	0.1020278870
8	0.43103755	5.609704190	0.356087266	-0.0249795719
10	-5.84421041	-5.256574708	-4.002253510	-0.1339474367
10	2.86525240	0.198942961	0.268941651	-0.0386373359
11	-4.56418450	-5.218905451	-4.949732979	0.0362845462
12	0.29141893	5.689588108	0.275202721	-0.0295591625
12	-0.74997226	-0.432703551	5.371038787	-0.0843607786
13	3.31046605	0.198775795	-0.109463620	0.0282992025
13	2.51079628	0.119879912	0.425872354	-0.0746824601
15	-5.07614171	-5.258477744	-4.560522942	-0.0330890155
15	-0.98116908	-0.428830071	5.596003377	-0.1195879897
16	2.51268863	0.035673473	0.470884455	-0.0763219265
18	2.28343568	-0.013728396	0.652349261	-0.1115901183
18	2.54999227	0.012570231	0.473903937	-0.0772628975
18	0.04311964	5.614186405	0.549617731	-0.0669614729
19	0.38152670	5.693127159	0.390159630	-0.0370417826
19	-0.92642108	-0.455463127	5.664441021	-0.1261706196
20	2.58972736	0.127763631	0.577978638	-0.0832847907
20	2.58972736	0.127763631	0.577978638	-0.0832847907
21	2.94731408	0.104462617	0.411791487	-0.0514971359
21	2.41420090	0.051865361	0.768682137	-0.1201515776
22	3.58154561	0.099212396	0.071955013	0.0136665063
24	3.61083877	0.088476912	0.069370265	0.0141215179
24	0.30429638	5.611197203	0.680420033	-0.0785587201
25	-5.61339787	-5.294528175	-3.836788369	-0.1508217402
25	2.82950835	0.134690867	0.612852116	-0.0898388603
25	3.89573471	0.239885377	-0.100929183	0.0474700232
27	3.58142015	0.035965106	-0.012486400	0.0100912921
29	3.08096319	-0.029320500	0.340561572	-0.0580965684
30	3.66517654	-0.002180881	-0.015996460	0.0099712727
30	-4.24461650	-5.378802667	-4.822527594	0.0176428345
31	4.07880706	-0.004136789	-0.309597422	0.0621515536
31	1.43865614	5.584330070	-0.144660966	0.0552893677
33	-5.26428865	-5.489590441	-4.144199477	-0.1174068998
35	0.31476272	5.414474635	0.491977377	-0.0817586371
36	1.09997292	5.648511891	-0.037288017	0.0193725861
42	-4.56383862	-5.290320753	-4.746382120	-0.0168500427
43	0.03082201	-0.448882534	5.021123707	-0.0346055389

44	-4.60329159	-5.345280096	-4.736965944	-0.0167937551
45	0.23232406	5.538574539	0.469607600	-0.0827275704
48	-0.58615518	5.631682635	1.010606707	-0.1877407583
49	-0.97064135	0.081016043	5.646482183	-0.1568068911
51	3.33244983	0.652080133	0.081397610	-0.0126800566
51	2.53278006	0.573184250	0.616733584	-0.1156617192
51	0.55902062	6.227397679	0.335556728	-0.0367058529
52	3.71672951	0.778055450	-0.088462609	0.0181920962
52	0.67674370	6.327074369	0.344141834	-0.0401609210
52	-0.09809090	0.231081337	5.261532575	-0.0606353162
53	-0.06969046	0.357553693	5.316145041	-0.0606087081
54	4.00215149	0.921152238	-0.151426797	0.0521836552
54	3.73559490	0.894853611	0.027018528	0.0178564343
56	4.07553782	0.915423815	-0.160170297	0.0536646467
59	2.77640587	0.781575704	0.727200729	-0.1171253418
61	3.94423895	0.868596221	0.033209499	0.0159395588
63	3.45801148	0.811905210	0.385865275	-0.0520353567
72	-4.37583414	-4.574090053	-4.418986342	-0.0449482658
73	0.97589079	6.346543814	0.415522966	-0.0424597177
80	3.49142684	0.915597488	0.342777986	-0.0531407271
80	0.45144104	6.464616406	0.775382430	-0.1114937442
82	-4.88082805	-4.165277209	-4.076156421	-0.1202246561
84	1.48136336	6.866796046	-0.010401130	0.0190533014
87	3.73745300	1.356093989	0.085449707	-0.0245509162
90	1.00971036	6.902161326	0.298557589	-0.0480921035
92	1.28842803	7.163512377	0.119025759	-0.0144673463
95	1.57281020	7.375797125	-0.065795206	0.0203682935
95	4.07968283	1.774180951	-0.141509000	0.0100668689
99	4.26349916	1.956966687	-0.169479960	0.0123913358
99	4.66333405	1.996414629	-0.437147947	0.0638821671
100	0.58282981	1.439727869	5.158867829	-0.0641489786
103	1.06264926	1.447050427	4.846658425	0.0040129419
105	1.15310587	1.469345801	4.889015271	0.0054979452
110	-3.46969369	-3.409080388	-4.855908291	0.0237788666
111	-3.77278346	-3.475214806	-4.722725380	-0.0100706374
111	0.58874576	1.380731495	5.255355614	-0.0622423908
112	-3.56436077	-3.500579193	-4.848824297	0.0237851798
117	4.57294703	1.859877054	-0.267973538	0.0509401158
117	1.79951782	7.435194600	-0.013814419	0.0269143196
118	-3.76159361	-3.365114457	-4.739469416	-0.0066923332
122	4.62213581	2.004369920	-0.356821522	0.0546108077
126	-3.79001040	-3.376711559	-4.640440394	-0.0427128971
132	2.05529236	7.562881864	-0.215156220	0.0283334244
133	1.19493099	1.681702095	4.772248453	-0.0071985722
139	4.88540840	2.225352749	-0.406217506	0.0549039660
140	1.94520678	7.800064585	0.008899847	-0.0013973528
143	1.76906884	2.110327452	4.552579730	0.0483520790
144	-4.44491857	-2.917509594	-4.122635328	-0.1387590335
151	3.94315965	2.284473938	-0.128622410	-0.0501703832
153	4.44719024	2.394095621	-0.340418714	-0.0172592147
156	1.37851705	1.970632020	4.614724475	-0.0266539373
162	2.43941155	8.116918512	-0.355563954	0.0282720677
162	1.66457695	2.020925481	4.561826787	0.0077976725
164	1.51121707	2.382467927	4.802283633	-0.0225110719
177	1.00408441	2.369916452	5.304741768	-0.0908926325
186	2.89005274	8.676021106	0.080384605	0.0644357290
200	1.94745248	2.895077498	5.165118313	0.0139587716

```

201      -2.64627747      -1.938656226      -4.499909626      0.0330510100
216      1.11327817      2.801347284      5.736130242     -0.0864638675
228     -3.21725594     -1.927062193     -3.823846559     -0.0385852016
231      1.50947285      2.849956284      5.620046734     -0.0224307538
242     -3.89592705     -2.212395930     -3.835492855     -0.0755741466
250      0.94602208      2.489442186      5.319001161     -0.0278590133
260      1.22532607      2.588703962      5.412748069      0.0066341720
278      0.71360379      2.596043093      5.981706411     -0.0604555614
283     -3.11469023     -2.086166176     -3.879133568      0.0578936239
287      4.16804560      3.158394911      1.181078571     -0.0162525124
314     -3.77455971     -1.980190257     -3.144650392     -0.0751143463
340      1.17266960      2.645621447      5.748034573      0.0052878497
357     -3.69327432     -2.169768857     -3.535215426     -0.0089394623
378      0.91357101      2.521848764      5.833785768      0.0089781190
384      3.94304469      3.087421165      1.410842110     -0.0123401480
389     -2.65458095     -1.769633346     -3.490788951      0.0684620722
392      4.02365703      3.358381242      1.842135966     -0.0716388443
411     -0.42411148     -0.013117745     -1.979559422     -0.0313414168
467     -0.12652659     -0.201782145     -2.814177635      0.0372223411
553      3.86296320      4.457746696      6.947506148     -0.0407081701
587      0.47170033      1.127078640      0.486109525     -0.0618688522
991      0.52544197      1.132380822      0.450132380     -0.0549479907
999      0.71214811      1.150801357      0.325142648     -0.0309039304
>

```

```
> vv<-cox.zph(mod4,transform='identity', terms=FALSE)
```

```
> vv
```

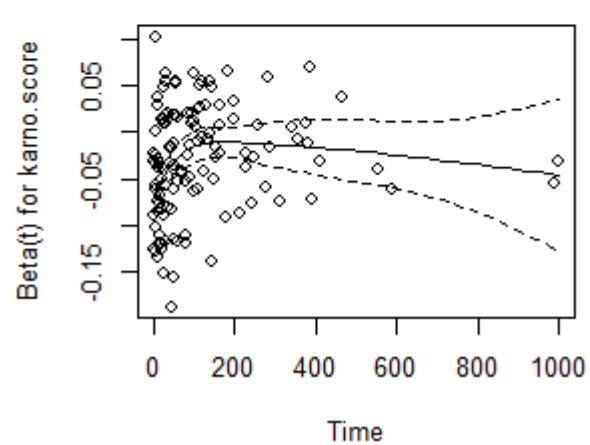
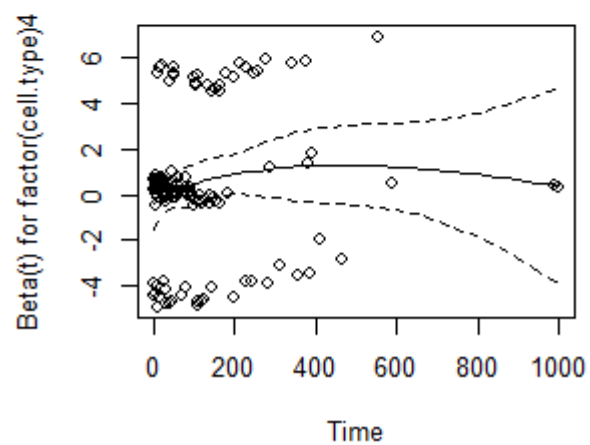
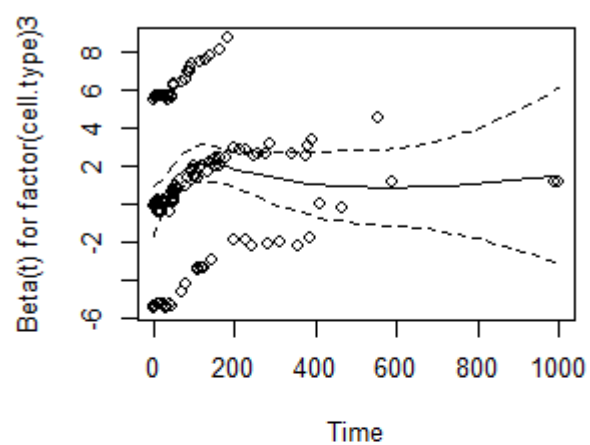
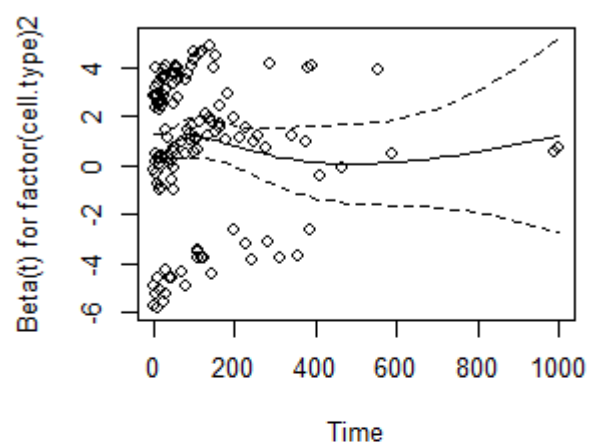
	chisq	df	p
factor(cell.type) 2	5.82	1	0.01580
factor(cell.type) 3	10.08	1	0.00150
factor(cell.type) 4	5.45	1	0.01959
karno.score	7.00	1	0.00814
GLOBAL	20.67	4	0.00037

```
> plot(vv,var=1)
```

```
>
```

```
> par(mfrow=c(2,2))
```

```
> plot(vv)
```



iv. Να κατασκευαστεί μια καμπύλη ROC για δύο διαφορετικούς χρόνους της επιλογής σας.

```
> mod4$linear.predictor
```

```
[1] -0.04421292 -0.35325223 -0.04421292 -0.04421292
[5] -0.35325223  1.19194429  0.57386569 -0.66229153
[9]  0.26482638 -0.35325223 -0.04421292  0.57386569
[13]  0.88290499 -0.66229153 -0.35325223  0.66793518
[17]  0.66793518  1.28601379  0.04985657  0.66793518
[21]  1.28601379  0.66793518  0.66793518  1.59505310
[25]  0.04985657  1.59505310  0.97697449  0.66793518
[29]  0.04985657  1.28601379  1.90409240  0.04985657
[33]  1.59505310  0.20437623  0.35889588  0.66793518
[37]  1.59505310  0.66793518  0.04985657  0.66793518
[41]  0.35889588  0.97697449  0.97697449  1.28601379
[45]  1.28601379  2.34274565  0.79754913  1.72466704
[49]  0.48850983  0.48850983  1.41562774  0.48850983
[53]  2.03370635  0.48850983  0.58996903 -0.33714888
[57]  0.58996903 -0.02810958  0.28092973  0.89900833
[61] -0.33714888 -0.33714888 -0.02810958 -0.64618819
[65] -0.64618819 -0.33714888 -0.33714888 -0.02810958
[69]  0.28092973 -0.97133084 -0.66229153 -0.66229153
[73]  0.26482638  0.26482638 -0.35325223 -0.35325223
[77]  1.19194429 -0.04421292 -0.97133084  0.88290499
[81]  1.19194429 -0.35325223 -0.97133084 -0.66229153
[85]  0.26482638 -0.35325223 -0.04421292 -0.97133084
[89]  0.26482638  1.59505310  0.35889588  1.90409240
[93]  1.59505310  0.66793518  1.28601379  1.59505310
[97]  1.90409240  0.66793518  0.35889588  0.04985657
[101] -0.10466308  0.35889588  0.35889588  0.35889588
[105]  0.97697449  1.59505310  1.28601379  1.72466704
[109]  1.72466704 -0.09866485  0.48850983  1.10658843
[113]  1.10658843  1.10658843  1.10658843  1.41562774
[117]  0.79754913  2.65178496  1.72466704  0.79754913
[121]  0.17947052  0.48850983  1.41562774  1.72466704
[125]  1.72466704  0.28092973 -0.02810958  1.20804764
[129]  0.28092973  1.20804764  0.28092973 -0.33714888
[133] -0.18262923  0.28092973 -0.02810958 -0.33714888
[137]  1.20804764
>
```

```
> library(risksetROC)
```

```
Loading required package: MASS
```

```
Attaching package: 'MASS'
```

```
The following object is masked from 'package:EnvStats':
```

```
boxcox
```

```
>
```

```
> eta<-mod4$linear.predictor
```

```
>
```

```
> ROC10=risksetROC(Stime= time, status=status, marker=eta, predict.time=60,
method="Cox", main="ROC Curve", lty=2, col="red", ylab="True Positive",
xlab="False Positive")
```

```
>
```

```
> ROC50=risksetROC(Stime= time, status=status,marker=eta, predict.time=80,
method="Cox", plot=FALSE)
```

```
>  
> lines(ROC50$FP,ROC50$TP, lty=3,col="darkblue")  
> legend(.6,.25,lty=c(2,3),col=c("red","darkblue"),  
+ legend=c("t=60","t=80"), bty="n")  
>
```

