

ANALISI DEL DATASET “PBC3”

A cura di Elia Mazzega e Luca Buratto

INTRODUZIONE

Il dataset analizza 349 pazienti affetti da *cirrosi biliare primitiva* (PBC), una malattia autoimmune del fegato, e provenienti da 6 diversi ospedali Europei. I pazienti si possono differenziare in due gruppi, ciascuno dei quali è stato sottoposto a un diverso trattamento: Ciclosporina A e placebo. Lo scopo dello studio è la determinazione dell'effetto di tali trattamenti sul tempo di sopravvivenza.

La prova è stata eseguita dal 1 gennaio 1983 fino al 1 gennaio 1987. Al momento dell'entrata dei pazienti nello studio, sono stati osservati diversi valori, di cui di seguito riportiamo la lista.

VARIABILI

- | | |
|-------------------|---|
| 1. <i>ptno</i> | identificazione del paziente |
| 2. <i>unit</i> | indica l'ospedale (1: Hvidovre, 2:Londra, 3:Copenhagen, 4: Barcellona, 5: Monaco di Baviera, 6:Lione) |
| 3. <i>tment</i> | trattamento (0:placeboM 1: Cya) |
| 4. <i>sex</i> | sexo (1:maschi; 0:femmine) |
| 5. <i>age</i> | età |
| 6. <i>stage</i> | stadio istologico (1,2,3,4) |
| 7. <i>gibleed</i> | precedente sanguinamento gastro-intestinale (1:si, 0:no) |
| 8. <i>crea</i> | creatina |
| 9. <i>alb</i> | albumina |
| 10. <i>bili</i> | bilirubina |
| 11. <i>alkph</i> | fosfatasi alcalina |
| 12. <i>asptr</i> | transaminasi aspartato |
| 13. <i>weight</i> | peso corporeo |
| 14. <i>days</i> | tempo di osservazione in giorni. Rappresenta il nostro follow-up |
| 15. <i>status</i> | lo stato di uscita (0:censurato, 1:fegato trapiantato, 2:morto) |

OPERAZIONI PRELIMINARI

Il trapianto di fegato è una delle indicazioni migliori per i pazienti con un livello istologico della malattia piuttosto alto. Tuttavia, poiché noi siamo interessati all'effetto del trattamento somministrato, considereremo il trapianto di fegato, assieme alla morte, come fallimento della prova.

- Generiamo dunque la variabile *fail*, che vale 1 se l'individuo è morto o ha subito il trapianto, e 0 se è stato censurato. La variabile viene generata dai seguenti comandi:

```
generate fail=0
replace fail=1 if status==1
replace fail=1 if status==2
```

- Per semplicità, raggruppiamo i pazienti in 4 classi di età, attraverso le seguenti istruzioni:

```
gen class_age=0
replace class_age =1 if age <=41
replace class_age =2 if 41< age & age<=52
replace class_age =3 if 52< age & age<=63
replace class_age =4 if 63< age & age<=75
```

- Cerchiamo ora di interpretare le variabili di tipo clinico che ci sono state fornite. Sappiamo da diversi studi scientifici che i dati normali di creatinina, albumina sierica, bilirubina, fosfatasi alcalina e transaminasi asportato sono i seguenti:

Sesso\variabili	Creatinina	Albumina sierica	Bilirubina	Fosfatasi alcalina	Transaminasi aspartato
Maschi	60-110 $\mu\text{mol/L}$	35-53 g/L	5-17 $\mu\text{mol/L}$	<170 UI/L	10-45 UI/L
Femmine	45-90 $\mu\text{mol/L}$	35-53 g/L	5-17 $\mu\text{mol/L}$	<170 UI/L	5-31 UI/L

Detto ciò, può essere conveniente creare 5 variabili tricotomiche, che assumono valore 0 se rientrano nel range di valori normali, 1 se sono sotto la norma e 2 se sono sopra. Il comando che si utilizza è il seguente

```
generate creanorm=0
replace creanorm=1 if crea<60 & sex==1
replace creanorm=2 if crea>110 & sex==1
replace creanorm=1 if crea<45 & sex==0
replace creanorm=2 if crea>90 & sex==0
```

Le altre variabili si creano con istruzioni molto simili. Si generano così le variabili *creanorm*, *albnorm*, *bilinorm*, *alkphnorm*, *asptrnorm*.

ANALISI DESCRITTIVA

1. Analisi monovariata delle variabili

Per l'analisi seguente utilizziamo il comando `summarize`. Le variabili *creanorm*, *albnorm*, *bilinorm*, *alkphnorm*, *asptrnorm* sono state considerate qui di seguito come dicotomiche (0=livello nella norma, 1=livello fuori norma).

variable	Obs	Mean	Std. Dev.	Min	Max
ptno	349	390.7393	168.5415	1	736
unit	349	3.037249	1.381857	1	6
tment	349	.504298	.5006994	0	1
sex	349	.1461318	.3537455	0	1
age	349	54.08883	9.943843	26	75
stage	291	2.683849	1.084235	1	4
gibleed	349	.1432665	.3508474	0	1
crea	342	77.98927	18.27808	35	199
alb	343	38.36627	5.662692	20.63	58
bili	349	45.4728	67.65207	2.333	453.1
alkph	349	996.5439	751.5989	66.33	5108
asptr	346	95.37908	52.91397	10.5	316.3
weight	339	60.34218	10.19663	38	98
days	349	942.6533	514.0021	1	2146
status	349	.4326648	.7724934	0	2
fail	349	.2578797	.4380955	0	1
creanorm	349	.1862464	.3898643	0	1
albnorm	349	.2464183	.4315437	0	1
bilinorm	349	.5902579	.4924921	0	1
alkphnorm	349	.982808	.1301728	0	1
asptrnorm	349	.9283668	.2582499	0	1
class_age	349	2.65616	.9171722	1	4

Osservazioni:

- Il gruppo di pazienti trattato con ciclosporina A ha numerosità pari alla metà del campione, l'altra metà è trattata col placebo.
- Come noto dalle ricerche sulla PBC, le donne sono le più affette da tale malattia. Nel nostro caso, il rapporto tra maschi e femmine è circa 1 a 15. L'età media all'entrata dello studio è di 54 anni (difatti la PBC colpisce le donne di mezz'età nel periodo prossimo alla menopausa).
- Circa un quarto dei pazienti ha fallito il trattamento medico (trapianto / morte).
- Il 59% dei pazienti presenta un livello di bilirubina fuori norma, mentre ciò che risulta ovvio è il fatto che la fosfatasi e la transaminasi sono sballate per la quasi totalità dei pazienti. Infatti la fosfatasi è uno dei principali fattori che, se

persistentemente molto alti, è segno di presenza della malattia, mentre la transaminasi è alta a causa del danneggiamento del fegato. Pertanto le escludiamo dal modello.

2. Analisi delle associazioni delle variabili

Consideriamo l'associazione più semplice per la nostra analisi, ossia quella che lega il TRATTAMENTO con il suo FALLIMENTO. La tabulazione risultante è la seguente.

```
tabulate tment fail, row col
```

tment	fail 0	1	Total
0	127 73.41 49.03	46 26.59 51.11	173 100.00 49.57
1	132 75.00 50.97	44 25.00 48.89	176 100.00 50.43
Total	259 74.21 100.00	90 25.79 100.00	349 100.00 100.00

Notiamo che le percentuali di fallimento in relazione al diverso trattamento sono pressoché le stesse. Si nota solo una leggerissima diminuzione dei fallimenti tra i pazienti curati con ciclosporina (51.11 contro 48.89), ma non è una differenza significativa. Proviamo ora ad inserire nello studio la variabile *gibleed*.

```
tabulate tment fail if gibleed==0, row col
```

tment	fail 0	1	Total
0	109 76.76 47.60	33 23.24 47.14	142 100.00 47.49
1	120 76.43 52.40	37 23.57 52.86	157 100.00 52.51
Total	229 76.59 100.00	70 23.41 100.00	299 100.00 100.00

```
tabulate tment fail if gibleed==1, row col
```

tment	fail 0	1	Total
0	18 58.06 60.00	13 41.94 65.00	31 100.00 62.00
1	12 63.16 40.00	7 36.84 35.00	19 100.00 38.00
Total	30 60.00 100.00	20 40.00 100.00	50 100.00 100.00

Si nota che, stratificando per *gibleed*, fra coloro che hanno avuto problemi di sanguinamento gastro-intestinale la proporzione dei fallimenti aumenta rispetto a coloro che non ne hanno avuti. Comunque, il diverso trattamento sembra essere irrilevante tra i pazienti che non hanno presentato questi problemi all'entrata nello studio. Tra quelli che invece li hanno presentati, notiamo che la cura con ciclosporina A ha ridotto lievemente la frequenza di fallimenti della prova rispetto a quelli trattati col placebo. In ogni caso, si nota che c'è stato un aumento dei fallimenti tra i pazienti curati con ciclosporina e aventi problemi di sanguinamento, rispetto a coloro curati nello stesso modo ma senza tali problemi. Non è un

gran risultato, però ci permette di affermare che la variabile *gibleed* ha un effetto interattivo tra TRATTAMENTO e FALLIMENTO.

Un'altra variabile che potrebbe interagire nello studio è il livello di creatina. Difatti, attraverso la seguente tabella, possiamo notare come la ciclosporina A abbia avuto pieno successo rispetto al trattamento con placebo nei pazienti con livello di creatinina più bassa della norma (*creanorm*==1).

```
tabulate tment fail if creanorm==1, row col
```

tment	fail 0	1	Total
0	0 0.00 0.00	5 100.00 100.00	5 100.00 50.00
1	5 100.00 100.00	0 0.00 0.00	5 100.00 50.00
Total	5 50.00 100.00	5 50.00 100.00	10 100.00 100.00

Come detto, abbiamo suddiviso i pazienti in 4 classi per età. Effettuando le tabulazioni per *class_age* si ottengono le seguenti tabelle:

```
tab tment fail if class_age ==1, row col
```

tment	fail 0	1	Total
0	20 80.00 52.63	5 20.00 45.45	25 100.00 51.02
1	18 75.00 47.37	6 25.00 54.55	24 100.00 48.98
Total	38 77.55 100.00	11 22.45 100.00	49 100.00 100.00

```
tab tment fail if class_age ==2, row col
```

tment	fail 0	1	Total
0	26 66.67 40.63	13 33.33 81.25	39 100.00 48.75
1	38 92.68 59.38	3 7.32 18.75	41 100.00 51.25
Total	64 80.00 100.00	16 20.00 100.00	80 100.00 100.00

```
tab tment fail if class_age ==3, row col
```

tment	fail 0	1	Total
0	61 74.39 53.04	21 25.61 44.68	82 100.00 50.62
1	54 67.50 46.96	26 32.50 55.32	80 100.00 49.38
Total	115 70.99 100.00	47 29.01 100.00	162 100.00 100.00

```
tab tment fail if class_age ==4, row col
```

tment	fail 0	1	Total
0	20 74.07 47.62	7 25.93 43.75	27 100.00 46.55
1	22 70.97 52.38	9 29.03 56.25	31 100.00 53.45
Total	42 72.41 100.00	16 27.59 100.00	58 100.00 100.00

Possiamo notare come la ciclosporina A abbia ridotto di molto i fallimenti nei pazienti di età compresa tra 42 e 52 anni: si ha una frequenza dell' 81% per quelli curati col placebo, 19% per quelli trattati col farmaco.

ANALISI NON PARAMETRICA DELLA SOPRAVVIVENZA

Quest'analisi ci permetterà di valutare la funzione di rischio e di sopravvivenza del campione. Qui la probabilità di sopravvivere è la probabilità di non subire l'evento "fegato trapiantato o decesso del paziente".

Creiamo la scala temporale tramite il seguente comando:

```
stset days, fail(fail) id(ptno)
```

Osserviamo che tale istruzione fornisce inoltre quattro nuove variabili, che però non risultano essere rilevanti per la nostra analisi (sono intrinseche in quelle già fornite).

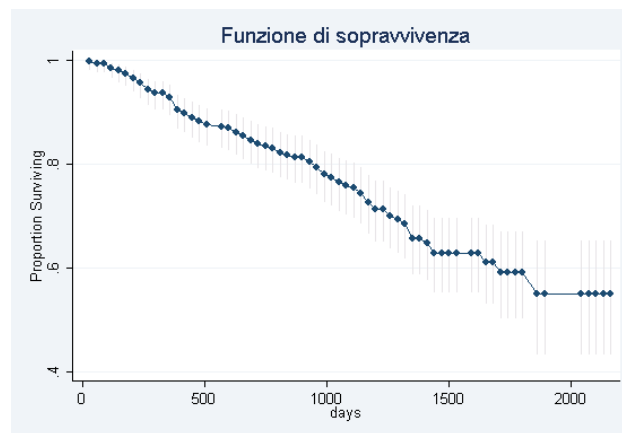
1. *Tabelle di sopravvivenza*

Sono state costruite le tabelle di sopravvivenza, di morte cumulata e di rischio tramite il comando

```
ltable days fail, intervals(365) su h f
```

Tuttavia una più facile interpretazione viene data dalla rappresentazione grafica della funzione di sopravvivenza, che si ottiene tramite il comando

```
ltable days fail, intervals(30) gr title("Funzione di sopravvivenza")
```

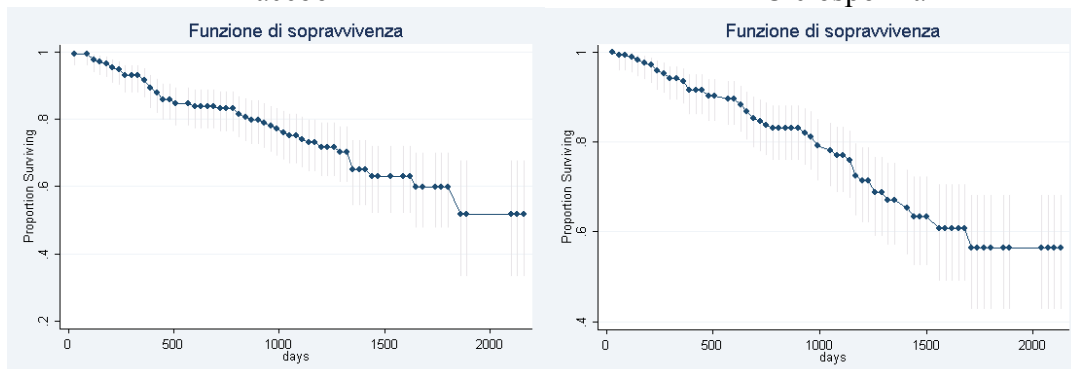


Alla fine dello studio i soggetti ancora vivi che non hanno subito trapianti sono circa il 55% dei 349 pazienti considerati. La pendenza poco accentuata (quasi lineare) è data dal fatto che la mortalità ha un effetto rilevante solo fino alla metà del quarto anno. Da lì in poi la lieve discesa della curva è imputabile alle poche censure rimaste.

Poiché il nostro scopo è valutare l'effetto del diverso trattamento utilizzato, distinguiamo le funzioni di sopravvivenza per gli individui che sono stati curati con la ciclosporina A e quelli trattati con placebo.

Placebo

Ciclosporina A



Tramite le tabelle di sopravvivenza dei due gruppi di pazienti, si nota un andamento pressoché uguale, con la un leggero aumento di probabilità di sopravvivenza per coloro che appartengono alle ultime classi fra i curati con ciclosporina A.

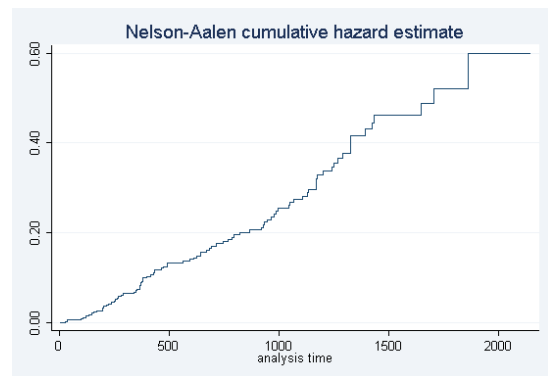
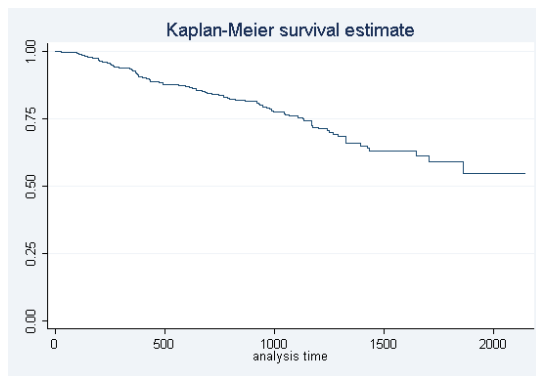
2. *Kaplan-Meier e Nelson-Aalen*

Possiamo utilizzare gli stimatori di Kaplan-Meier e di Nelson-Aalen per stimare in un altro modo la funzione di sopravvivenza e rischio cumulato. Computeremo rispettivamente

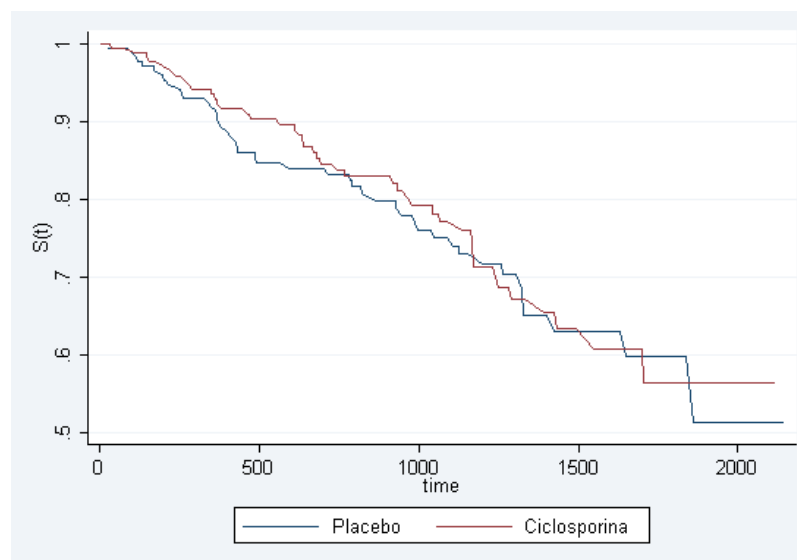
```
sts graph
```

e

```
sts graph, na
```



Differenziando l'analisi per il diverso trattamento, sovrapponiamo quindi i grafici delle due stime ottenute con Kaplan-Meier, che riconferma l'analisi eseguita al punto precedente. Osserviamo come la curva dei pazienti trattati con ciclosporina domini lievemente l'altra.



Per un confronto più accurato delle due curve, eseguiamo il *log-rank test*, attraverso le seguenti istruzioni.

```
sts test tment
```

```
      failure _d: fail
analysis time _t: days
           id: ptno
```

Log-rank test for equality of survivor functions

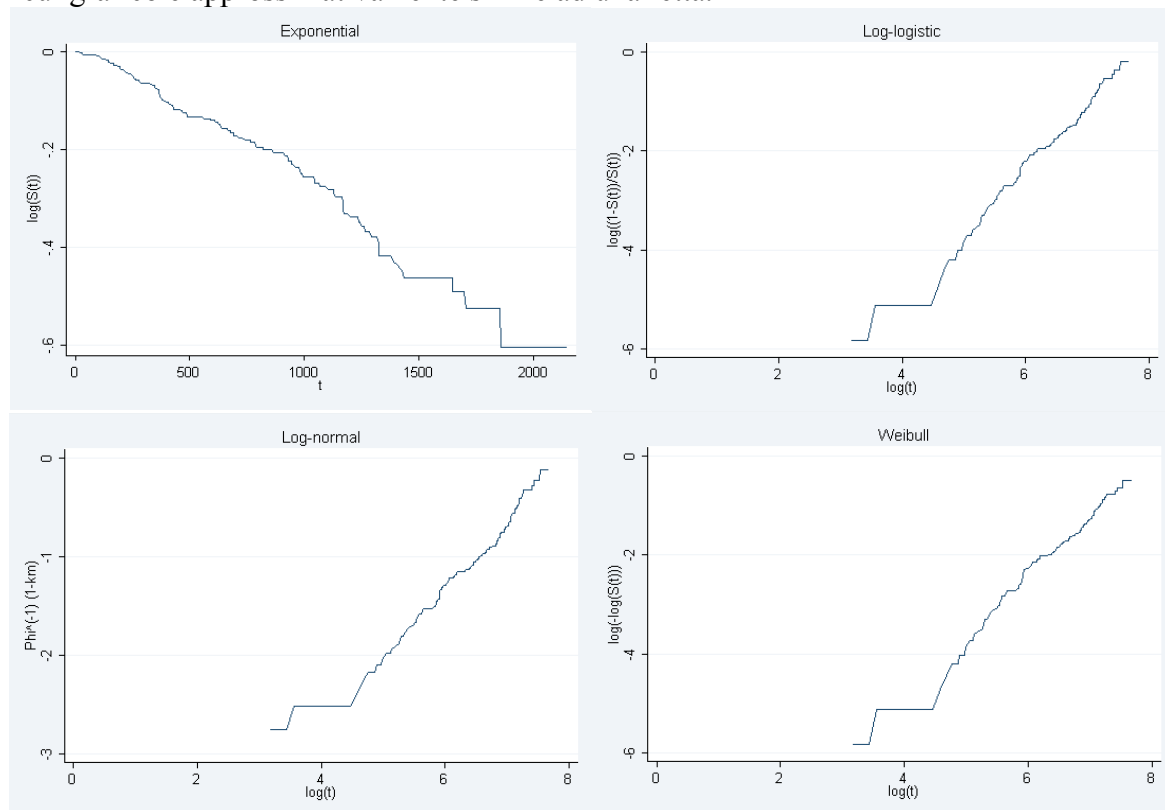
tment	Events observed	Events expected
0	46	44.68
1	44	45.32
Total	90	90.00
chi2(1) =		0.08
Pr>chi2 =		0.7813

Si evince dall'analisi che l'andamento della curva di sopravvivenza nei due gruppi è significativamente (95%) uguale.

ANALISI PARAMETRICA

1. Analisi preliminare

Per scegliere il modello parametrico più appropriato, si effettuano delle trasformazioni della funzione di sopravvivenza ottenuta dalla stima di Kaplan-Meier. Scegliremo dunque il modello il cui grafico è approssimativamente simile ad una retta.



- Esponenziale: è la trasformazione logaritmica di $\hat{S}(t)$, ovvero della funzione di sopravvivenza di K-M;
- Log-Logistica: sull'asse x inseriamo il logaritmo del tempo, mentre sull'asse y la trasformazione $\ln\left(\frac{1-\hat{S}(t)}{\hat{S}(t)}\right)$;
- Log-Normale:tenendo sempre il logaritmo del tempo si effettua la trasformazione $\Phi^{-1}(1 - \hat{S}(t))$;
- Weibull: in ascissa resta il logaritmo del tempo, ed in ordinata quest'ultima trasformazione: $\ln(-\ln(1 - \hat{S}(t)))$;

Dall'analisi dei grafici, notiamo come tutte le curve approssimano piuttosto bene una retta. Tuttavia, in questo elaborato, considereremo il modello Log-Normale.

2. Analisi del modello scelto

• *Modello senza covariate*

Il comando per stimare la funzione di rischio attraverso il modello Log-Normale è il seguente:

```
streg, dist(ln)
```

```

failure _d: fail
analysis time _t: days
id: ptno

Iteration 0: log likelihood = -8568.1821 (not concave)
Iteration 1: log likelihood = -430.05553 (not concave)
Iteration 2: log likelihood = -271.22548
Iteration 3: log likelihood = -263.4534
Iteration 4: log likelihood = -263.3821
Iteration 5: log likelihood = -263.38206
Iteration 6: log likelihood = -263.38206

Lognormal regression -- accelerated failure-time form

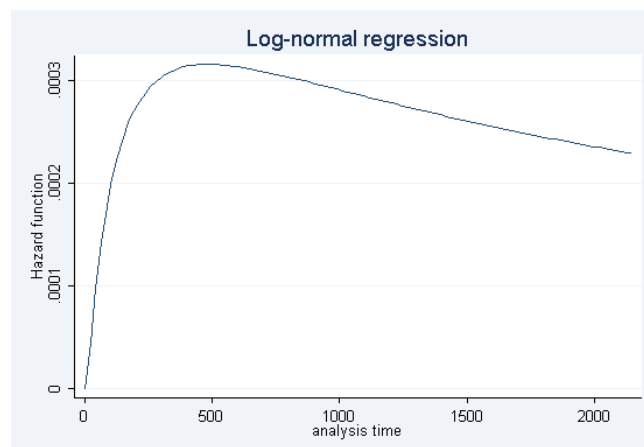
No. of subjects = 349          Number of obs = 349
No. of failures = 90
Time at risk = 328986

Log likelihood = -263.38206      wald chi2(0) = .
                                Prob > chi2 = .

+-----+-----+-----+-----+-----+-----+
|      _t      | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
+-----+-----+-----+-----+-----+-----+
|      _cons    | 7.89683 | .1432738 | 55.12 | 0.000 | 7.616018  8.177641 |
+-----+-----+-----+-----+-----+-----+
|      /ln_sig   | .3522377 | .0820443 | 4.29  | 0.000 | .1914339  .5130416 |
+-----+-----+-----+-----+-----+-----+
|      sigma     | 1.422247 | .1166872 |       |       | 1.210985  1.670364 |
+-----+-----+-----+-----+-----+-----+

```

I coefficienti del modello Log-Normale a e b sono dati pertanto da $a=7.89683$ e $b=1.4222$, che risultano essere significativamente diversi da 0 a livello del 95%. Il grafico della funzione di rischio si ottiene con l'istruzione `stcurve, hazard`



Poiché $b>1$, la curva segue un andamento di tipo campanulare.

- **Modello con covariate**

Valutiamo dapprima l'efficienza dei modelli con una sola covariata.

Modello con *tment*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_tment_1	.1166276	.2001887	0.58	0.560	-.275735	.5089903
_cons	7.837336	.1724856	45.44	0.000	7.49927	8.175401
/ln_sig	.3506283	.0820365	4.27	0.000	.1898398	.5114168
sigma	1.419959	.1164884			1.209056	1.667652

Dal coefficiente 0.11, si evince che il rischio di fallimento per i pazienti trattati con ciclosporina diminuisce rispetto ai non trattati.

Modello con *gibleed*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<i>_igibleed_1</i>	-.6941843	.2611615	-2.66	0.008	-1.206051	-.1823171
<i>_cons</i>	7.993032	.1524922	52.42	0.000	7.694152	8.291911
<i>/ln_sig</i>	.3349201	.0817253	4.10	0.000	.1747414	.4950988
<i>sigma</i>	1.397829	.114238			1.190938	1.64066

La variabile *gibleed* risulta essere significativa, e un coefficiente pari a -0.69 indica che il rischio aumenta per coloro che presentano sanguinamento gastro-intestinale.

Modello con *sex*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<i>_isex_1</i>	-.5769669	.2631013	-2.19	0.028	-1.092636	-.0612978
<i>_cons</i>	7.989705	.155005	51.54	0.000	7.685901	8.29351
<i>/ln_sig</i>	.3464676	.081961	4.23	0.000	.1858271	.5071082
<i>sigma</i>	1.414064	.115898			1.204214	1.660482

Anche questa variabile è significativa. Ciò è evidente poiché il sesso incide molto sul fallimento della prova, dato che la PBC colpisce di più le donne rispetto agli uomini.

Modello con *age*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<i>_iclass_ag~2</i>	.0904147	.3523167	0.26	0.797	-.6001132	.7809427
<i>_iclass_ag~3</i>	-.1760983	.3096252	-0.57	0.570	-.7829525	.4307559
<i>_iclass_ag~4</i>	.0625391	.3676373	0.17	0.865	-.6580169	.783095
<i>_cons</i>	7.949982	.2961315	26.85	0.000	7.369575	8.530389
<i>/ln_sig</i>	.3498333	.0821214	4.26	0.000	.1888782	.5107884
<i>sigma</i>	1.418831	.1165165			1.207894	1.666605

Si ha che il rischio aumenta per i pazienti di età compresa tra 53 e 63 anni.

Modello con *albnorm*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<i>_ialbnorm_1</i>	-.874172	.2194602	-3.98	0.000	-1.304306	-.4440378
<i>_ialbnorm_2</i>	-.4980876	.5687305	-0.88	0.381	-1.612779	.6166036
<i>_cons</i>	8.084412	.1588177	50.90	0.000	7.773135	8.395689
<i>/ln_sig</i>	.3043032	.0816063	3.73	0.000	.1443578	.4642486
<i>sigma</i>	1.35568	.110632			1.155297	1.590818

I pazienti con livelli di albumina fuori norma presentano un rischio di fallimento più elevato.

Modello con *bilinorm*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
<i>_ibilinorm_1</i>	-.5890837	.7836335	-0.75	0.452	-2.124977	.9468098
<i>_ibilinorm_2</i>	-1.783694	.2586997	-6.89	0.000	-2.290736	-1.276652
<i>_cons</i>	9.016445	.2675778	33.70	0.000	8.492003	9.540888
<i>/ln_sig</i>	.2160291	.079854	2.71	0.007	.0595182	.37254
<i>sigma</i>	1.241138	.0991098			1.061325	1.451417

Come prima, aumenta il rischio se il livello di bilirubina è sfasato. Teniamo dunque in considerazione i livelli dell'albumina e della bilirubina, provando anche un modello con queste due variabili, includendo ovviamente anche il trattamento.

Modello con *tment*, *albnorm* e *bilinorm*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Itment_1	.2520679	.1873278	1.35	0.178	-.1150879	.6192237
_Ialbnorm_1	-.4695487	.2103864	-2.23	0.026	-.8818984	-.0571989
_Ialbnorm_2	-.1455123	.537236	-0.27	0.787	-1.198475	.9074509
_Ibilinorm_1	-.4717893	.7662746	-0.62	0.538	-1.97366	1.030081
_Ibilinorm_2	-1.636054	.2563159	-6.38	0.000	-2.138424	-1.133684
_cons	8.889869	.2717463	32.71	0.000	8.357257	9.422482
/ln_sig	.1909503	.0797912	2.39	0.017	.0345624	.3473382
sigma	1.210399	.0965792			1.035167	1.415295

Costruiamo il modello con *tment*, *sex*, *gibleed*, *class_age*, *albnorm* e *bilinorm*.

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Itment_1	.2089654	.190589	1.10	0.273	-.1645823	.582513
_Iclass_ag~2	.1711137	.3229629	0.53	0.596	-.461882	.8041094
_Iclass_ag~3	-.1902579	.2853439	-0.67	0.505	-.7495217	.3690059
_Iclass_ag~4	-.0479832	.3444047	-0.14	0.889	-.723004	.6270376
_Isex_1	-.6394228	.2449295	-2.61	0.009	-1.119476	-.1593698
_Ialbnorm_1	-.3680147	.217248	-1.69	0.090	-.793813	.0577837
_Ialbnorm_2	-.3835358	.5442128	-0.70	0.481	-1.450173	.6831017
_Ibilinorm_1	-.3656466	.8005608	-0.46	0.648	-1.934717	1.203424
_Ibilinorm_2	-1.689512	.2640615	-6.40	0.000	-2.207063	-1.171961
_Igibleed_1	-.3555322	.2421787	-1.47	0.142	-.8301937	.1191294
_cons	9.155996	.3850188	23.78	0.000	8.401373	9.910619
/ln_sig	.1750985	.079664	2.20	0.028	.01896	.331237
sigma	1.191364	.0949087			1.019141	1.39269

Notiamo come alcune variabili in questo caso perdano di significatività sopra il 5%, ma compiendo una piccola forzatura possiamo comunque prenderle come buone.

Proviamo infine un modello prendendo in considerazione solo *tment* e *gibleed*:

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_Itment_1	.0532463	.1993513	0.27	0.789	-.3374751	.4439677
_Igibleed_1	-.6853813	.2630473	-2.61	0.009	-1.200944	-.1698182
_cons	7.964601	.1843598	43.20	0.000	7.603262	8.325939
/ln_sig	.3344292	.0817272	4.09	0.000	.1742469	.4946115
sigma	1.397143	.1141845			1.190349	1.639861

Anche in questo caso, la variabile *tment* perde di significatività, ma poiché essenziale per il nostro esperimento prendiamo comunque in considerazione anche questo modello. Effettuiamo il test del rapporto di verosimiglianza per comparare alcuni dei modelli migliori trovati finora.

Modello senza covariate

→ modello1

Modello con *tment*, *albnorm* e *bilinorm*

→ modello2

Modello con *tment*, *sex*, *gibleed*, *class_age*, *albnorm* e *bilinorm*

→ modello3

Modello con *tment* e *gibleed*

→ modello4.

Likelihood-ratio test
(Assumption: modello1 nested in modello4)

LR chi2(2) = 7.07
Prob > chi2 = 0.0291

Likelihood-ratio test
(Assumption: modello4 nested in modello3)

LR chi2(8) = 79.15
Prob > chi2 = 0.0000

Likelihood-ratio test
(Assumption: modello2 nested in modello3)

LR chi2(5) = 11.71
Prob > chi2 = 0.0390

Ne consegue che i modelli da noi scelti sono significativamente diversi tra loro. Per scegliere il modello più adeguato, utilizziamo il criterio di Akaike, attraverso la seguente istruzione

```
estat ic
```

modello1

Model	Obs	ll(nu11)	ll(model)	df	AIC	BIC
.	349	.	-263.3821	2	530.7641	538.4743

modello2

Model	Obs	ll(nu11)	ll(model)	df	AIC	BIC
.	349	-263.3821	-226.1242	7	466.2484	493.2339

modello3

Model	Obs	ll(nu11)	ll(model)	df	AIC	BIC
.	349	-263.3821	-220.2699	12	464.5398	510.8007

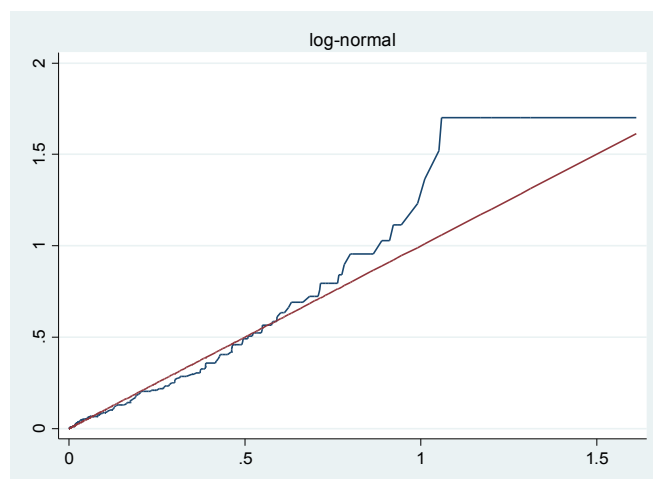
modello4

Model	Obs	ll(nu11)	ll(model)	df	AIC	BIC
.	349	-263.3821	-259.8464	4	527.6929	543.1132

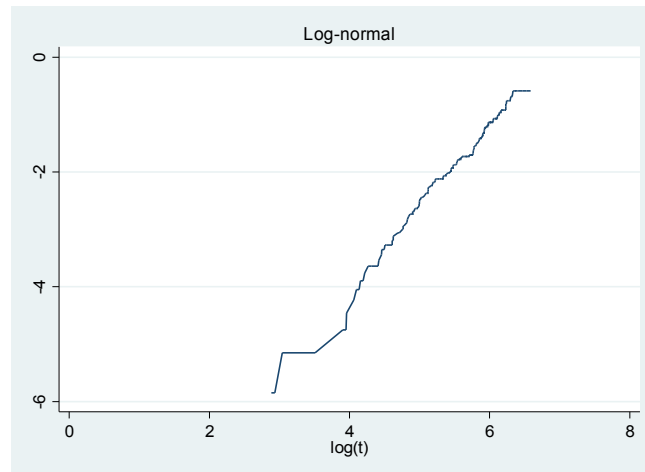
Il modello che noi andremo a scegliere sarà quello con statistica AIC più bassa, ovvero quello con *tment*, *sex*, *gibleed*, *class_age*, *albnorm* e *bilinorm*.

- **Analisi dei residui**

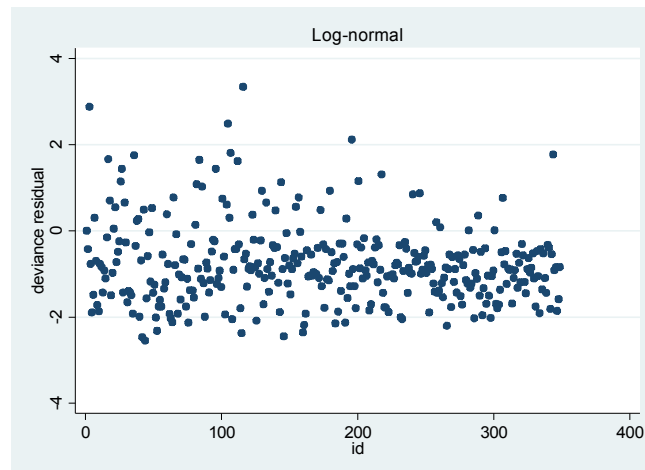
Analizziamo i residui di Cox-Snell ottenuti col modello scelto con le 6 covariate.



Poiché il grafico segue abbastanza bene la bisettrice del primo quadrante (almeno fino al valore 1), ne consegue che l'ipotesi di modello log-normale è corretta. Calcoliamo ora i residui standardizzati.



I residui approssimano abbastanza bene una retta. Infine, computiamo i residui di devianza.



I residui si distribuiscono in ordine piuttosto sparso. E' presente comunque una nuvola di punti concentrati attorno al valore nullo della varianza residua.

MODELLO DI COX

Utilizziamo il modello di Cox per calcolare i vari *Hazard ratio* delle variabili più rilevanti nello studio. La tabella si ottiene mediante la seguente istruzione.

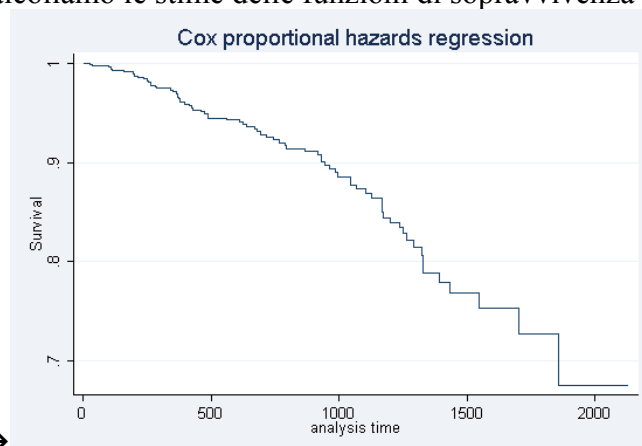
```
xi:stcox i.sex i.tment i.gibleed stage i.class_age i.creanorm i.albnorm
i.bilinorm weight unit
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
_Isex_1	3.871063	1.354884	3.87	0.000	1.949428	7.686939
_Itment_1	.9174519	.2209528	-0.36	0.721	.5722499	1.470892
_Igibleed_1	.9289627	.3046727	-0.22	0.822	.4884599	1.76672
stage	2.029695	.3250733	4.42	0.000	1.482871	2.778165
_Iclass_age~2	.4893894	.2202533	-1.59	0.112	.2025653	1.182344
_Iclass_age~3	1.161253	.4290069	0.40	0.686	.5629417	2.395466
_Iclass_age~4	.4218222	.2144231	-1.70	0.089	.1557549	1.142397
_Icreanorm_1	1.681086	.9197087	0.95	0.342	.575313	4.912194
_Icreanorm_2	1.599792	.6571624	1.14	0.253	.7151662	3.578656
_Ialbnorm_1	1.426117	.3992734	1.27	0.205	.8238385	2.468699
_Ialbnorm_2	3.230388	2.126367	1.78	0.075	.8891207	11.73677
_Ibilinorm_1	5.76e-16	5.91e-08	-0.00	1.000	0	.
_Ibilinorm_2	5.737322	2.448979	4.09	0.000	2.485269	13.24479
weight	.9618077	.0150653	-2.49	0.013	.9327288	.9917931
unit	.9858017	.1076679	-0.13	0.896	.7958341	1.221115

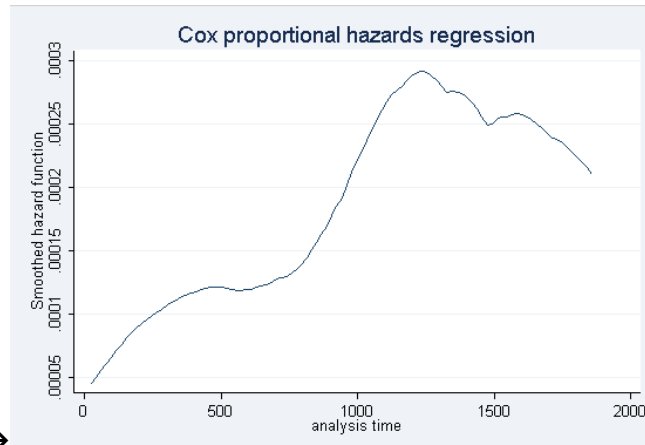
Alcune osservazioni:

- Si nota come i maschi abbiano una probabilità di fallimento quasi 4 volte maggiore rispetto alle femmine.
- I pazienti con livelli di creatina e albumina fuori norma hanno maggiore probabilità di fallire rispetto ai pazienti con valori normali.
- Il trattamento con ciclosporina abbassa la probabilità di fallimento di circa il 10% rispetto all'uso del placebo nell'esperimento.
- La presenza di valori anomali nella bilirubina aumenta di 6 volte il rischio di fallimento.

Tramite il modello di Cox, calcoliamo le stime delle funzioni di sopravvivenza e di rischio:

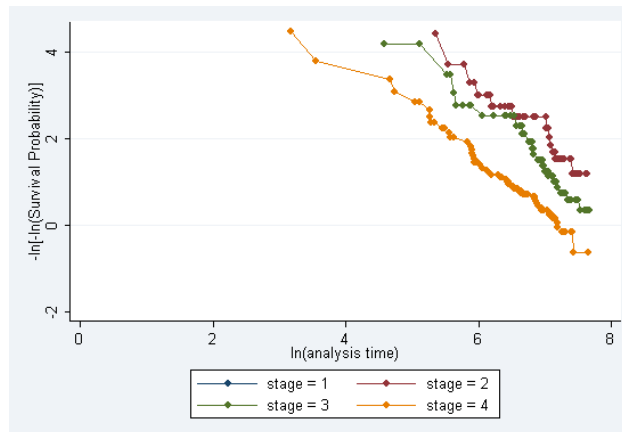


Funzione di sopravvivenza →

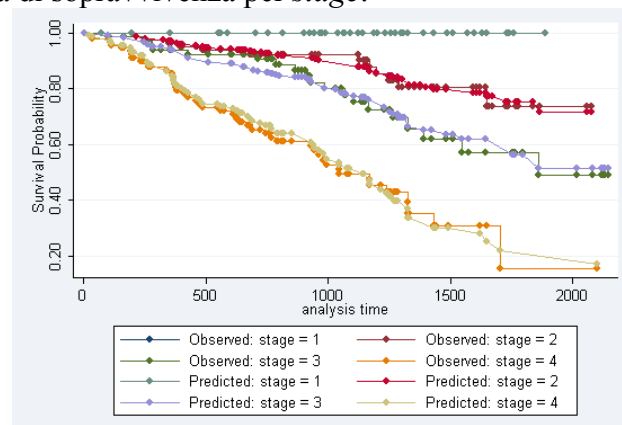


Funzione di rischio →

Verifichiamo infine se è rispettata la proporzionalità del modello di Cox, stratificando ad esempio la variabile *stage*.



Controlliamo ora le curve di sopravvivenza per *stage*.



L'ipotesi di proporzionalità è confermata dal fatto che dal primo grafico le curve sembrano essere abbastanza parallele tra loro, mentre dal secondo si evince la sovrapposizione fra funzioni di sopravvivenza empiriche e teoriche. Per verificare l'ipotesi di proporzionalità, in alternativa è possibile effettuare il test di Shoenfield:

Test of proportional-hazards assumption

Time: Time

	chi2	df	Prob>chi2
global test	17.47	15	0.2918

Come ci si aspettava, il test conferma l'ipotesi di proporzionalità del modello di Cox.

CONCLUSIONI

Lo scopo della prova era lo studio dell'effetto del trattamento con ciclosporina A. Tuttavia come abbiamo potuto constatare finora, questa cura non sembra aver portato a nessun risultato significativo, se non tra i pazienti con precedente sanguinamento gastro-intestinale, tra quelli con livelli di creatina sotto la norma e tra i pazienti in età compresa fra 42 e 52 anni. Nel procedimento non parametrico il trattamento con ciclosporina A produce in media un miglioramento sui pazienti, almeno nei primi 2 anni. Anche nel resto dell'elaborato si è potuta notare una leggera diminuzione del rischio di fallimento per i pazienti trattati con ciclosporina.