

**ANALISI DELLA SOPRAVVIVENZA:
IL MODELLO DI COX (parte I)**

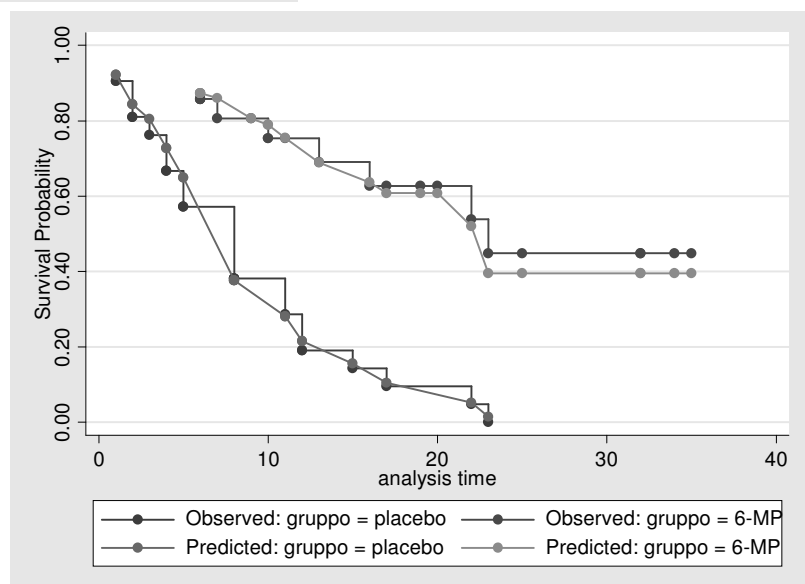
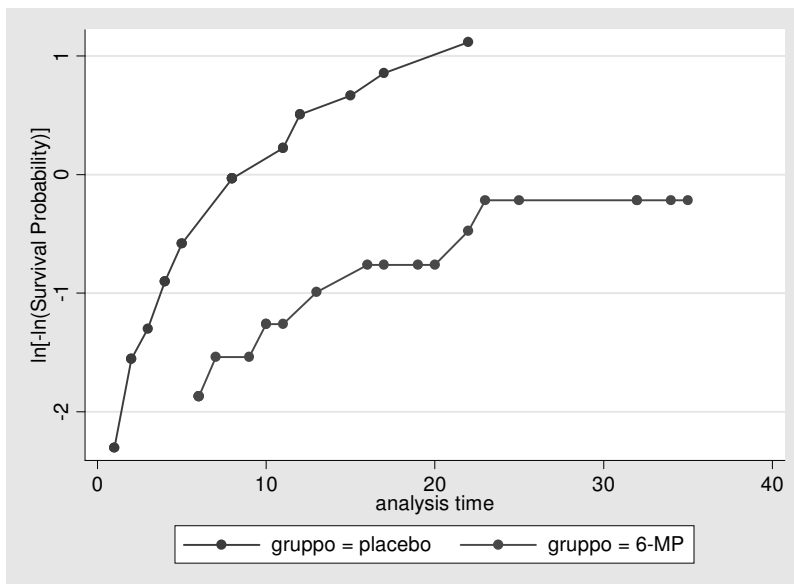
Il modello di Cox si basa sull'**assunzione di proporzionalità dei rischi**, che deve essere verificata affinché i risultati ottenuti siano validi. Con modelli che contengono un numero ridotto di covariate, è possibile utilizzare **metodi grafici**.

Costruire il grafico log(-log):

stphplot, by(gruppo) noltime nonegative

Confrontare la funzione di sopravvivenza stimata tramite la procedura di Kaplan-Meier con la curva predetta tramite il modello di Cox:

stcoxkm, by(gruppo)



**ANALISI DELLA SOPRAVVIVENZA:
IL MODELLO DI COX (parte II)**

ESEMPIO 2:

I dati sono riportati nel libro "E. Marubini, M.G. Valsecchi. Analysing survival data from clinical trials and observational studies. pag. 237" e si riferiscono ad una sperimentazione randomizzata su pazienti con cancro avanzato alle ovaie.

I dati sono forniti su floppy disk (nome file: **ovarian.dta**).

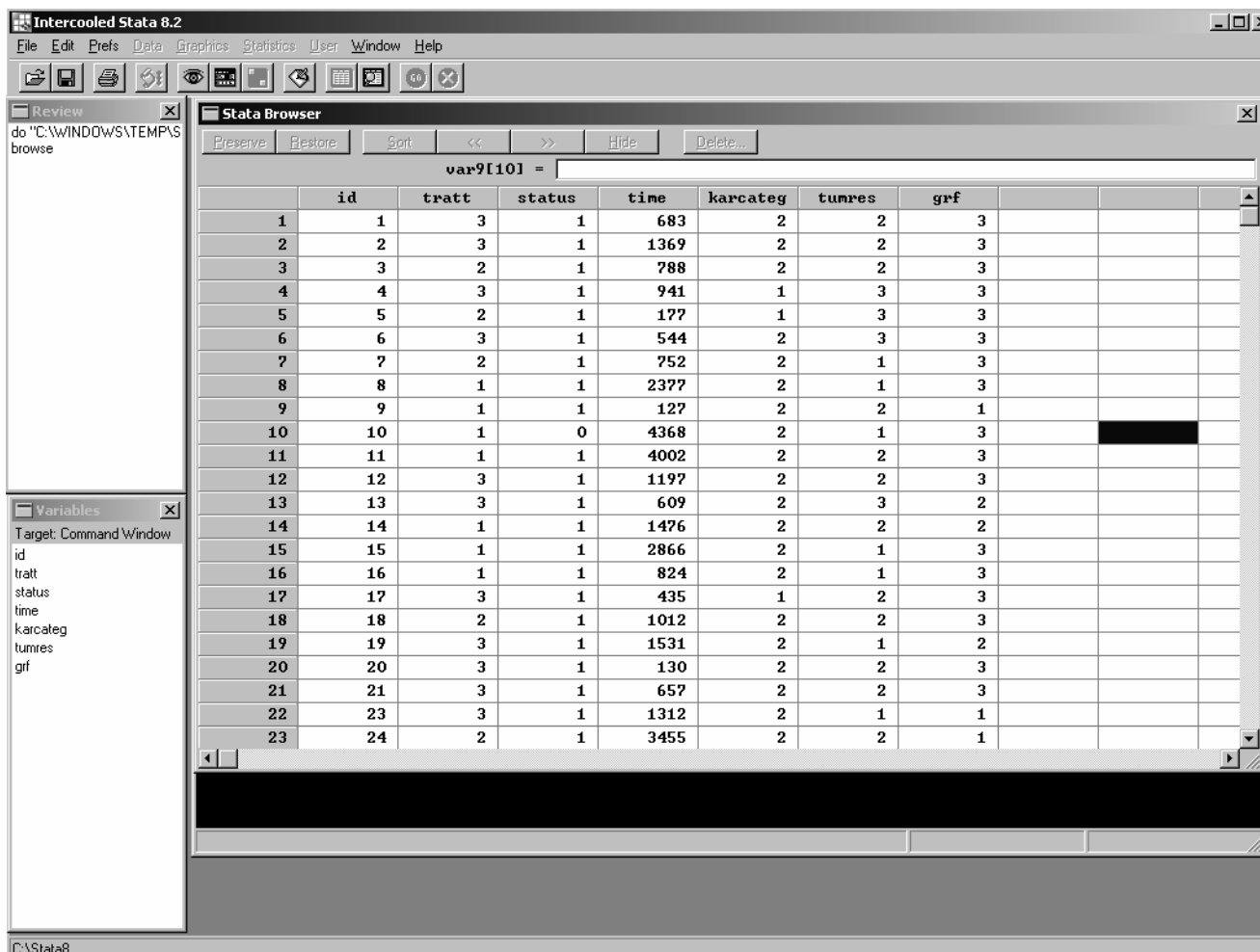
Il contenuto del file è il seguente:

variabili: codice identificativo (variabile '**id**')
 gruppo di terapia (variabile '**tratt**')
 1 = CAP = cyclophosphamide + adriamycin + platinum
 2 = CP = cyclophosphamide + platinum
 3 = P = platinum
 status (variabile '**status**')
 0 = censored
 1 = dead
 tempo di sopravvivenza in giorni (variabile '**time**')
 indice di Karnowsky (variabile '**karcateg**')
 1 = '≤70'
 2 = '>70'
 dimensione residuale del tumore (variabile '**tumres**')
 1 = '≤2 cm'
 2 = '(2-10] cm'
 3 = '>10 cm'
 grado Figo (variabile '**grf**')

numero casi: **476**

Caricare in memoria un file di dati nel formato STATA (.dta):

use a:\ovarian.dta



The screenshot shows the Stata 8.2 interface. The main window is titled "Stata Browser" and displays a data table with the following columns: id, tratt, status, time, karcateg, tumres, and grf. The data is as follows:

| | id | tratt | status | time | karcateg | tumres | grf |
|----|----|-------|--------|------|----------|--------|-----|
| 1 | 1 | 3 | 1 | 683 | 2 | 2 | 3 |
| 2 | 2 | 3 | 1 | 1369 | 2 | 2 | 3 |
| 3 | 3 | 2 | 1 | 788 | 2 | 2 | 3 |
| 4 | 4 | 3 | 1 | 941 | 1 | 3 | 3 |
| 5 | 5 | 2 | 1 | 177 | 1 | 3 | 3 |
| 6 | 6 | 3 | 1 | 544 | 2 | 3 | 3 |
| 7 | 7 | 2 | 1 | 752 | 2 | 1 | 3 |
| 8 | 8 | 1 | 1 | 2377 | 2 | 1 | 3 |
| 9 | 9 | 1 | 1 | 127 | 2 | 2 | 1 |
| 10 | 10 | 1 | 0 | 4368 | 2 | 1 | 3 |
| 11 | 11 | 1 | 1 | 4002 | 2 | 2 | 3 |
| 12 | 12 | 3 | 1 | 1197 | 2 | 2 | 3 |
| 13 | 13 | 3 | 1 | 609 | 2 | 3 | 2 |
| 14 | 14 | 1 | 1 | 1476 | 2 | 2 | 2 |
| 15 | 15 | 1 | 1 | 2866 | 2 | 1 | 3 |
| 16 | 16 | 1 | 1 | 824 | 2 | 1 | 3 |
| 17 | 17 | 3 | 1 | 435 | 1 | 2 | 3 |
| 18 | 18 | 2 | 1 | 1012 | 2 | 2 | 3 |
| 19 | 19 | 3 | 1 | 1531 | 2 | 1 | 2 |
| 20 | 20 | 3 | 1 | 130 | 2 | 2 | 3 |
| 21 | 21 | 3 | 1 | 657 | 2 | 2 | 3 |
| 22 | 23 | 3 | 1 | 1312 | 2 | 1 | 1 |
| 23 | 24 | 2 | 1 | 3455 | 2 | 2 | 1 |

Assegnare label ai valori della variabile gruppo di terapia:

```
label define tr_code 1 "CAP" 2 "CP" 3 "P"  
label values tratt tr_code
```

Definire i dati di sopravvivenza in STATA:

```
stset time, failure(status==1)
```

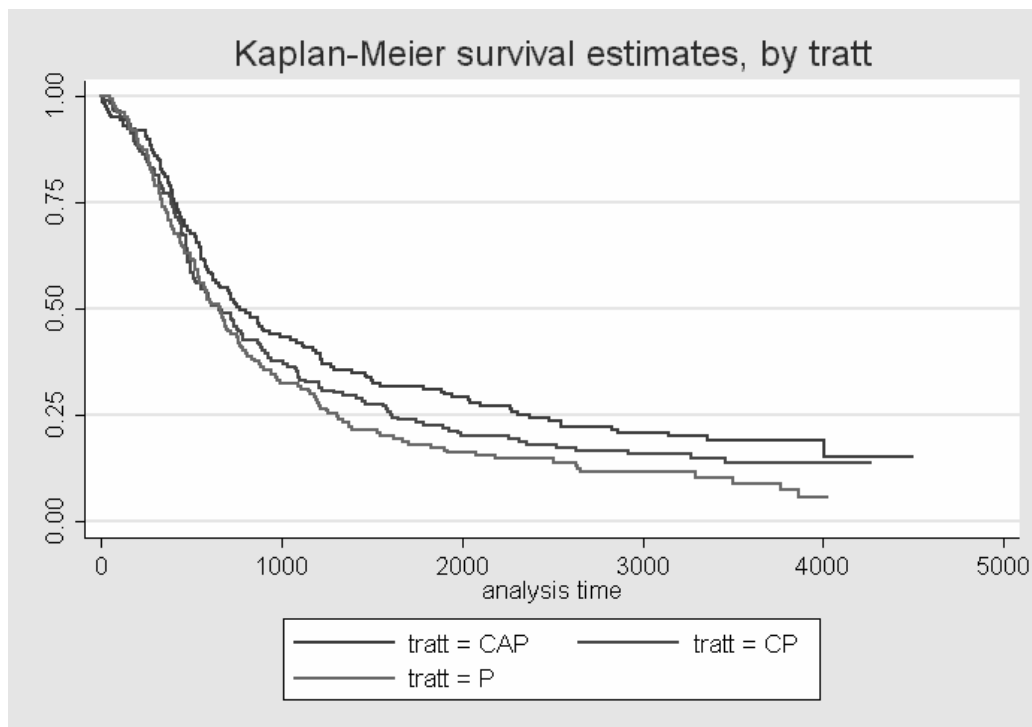
```
failure event:  status == 1  
obs. time interval:  (0, time]  
exit on or before:  failure
```

```
-----  
476 total obs.  
0 exclusions  
-----
```

```
476 obs. remaining, representing  
403 failures in single record/single failure data  
561396 total analysis time at risk, at risk from t = 0  
earliest observed entry t = 0  
last observed exit t = 4493
```

Tracciare il grafico della funzione di sopravvivenza stimata tramite la procedura di Kaplan-Meier:

```
sts graph, by(tratt)
```



Calcolare il tempo di sopravvivenza mediano:

stci, by(tratt) median

```
failure _d: status == 1  
analysis time _t: time
```

| tratt | no. of subjects | 50% | Std. Err. | [95% Conf. Interval] |
|-------|-----------------|-----|-----------|----------------------|
| CAP | 157 | 767 | 51.94369 | 615 1084 |
| CP | 162 | 653 | 45.32683 | 508 868 |
| P | 157 | 657 | 30.81726 | 536 763 |
| total | 476 | 689 | 41.93972 | 594 767 |

Confrontare la sopravvivenza tra gruppi diversi:

sts test tratt

```
failure _d: status == 1  
analysis time _t: time
```

Log-rank test for equality of survivor functions

| tratt | Events observed | Events expected |
|-------|-----------------|-----------------|
| CAP | 126 | 147.92 |
| CP | 136 | 134.08 |
| P | 141 | 121.00 |
| Total | 403 | 403.00 |

```
chi2(2) = 6.63  
Pr>chi2 = 0.0363
```

Poiché il modello di Cox dovrà contenere **covariate categoriali** (gruppo di terapia, indice di Karnowsky, dimensione residuale del tumore, grado Figo), è necessario specificare che i valori delle variabili sono codici.

Costruire il modello di Cox con più covariate categoriali:

xi: stcox i.tratt i.karcateg i.tumres i.grf

The screenshot shows the Stata 8.2 interface with the following content:

Review

```
do "C:\WINDOWS\TEMPAS
browse
do "C:\WINDOWS\TEMPAS
do "C:\WINDOWS\TEMPAS
do "C:\WINDOWS\TEMPAS
do "C:\WINDOWS\TEMPAS
do "C:\WINDOWS\TEMPAS
do "C:\WINDOWS\TEMPAS
```

Stata Results

```
. do "C:\WINDOWS\TEMP\STD00000000.tmp"
. xi: stcox i.tratt i.karcateg i.tumres i.grf
i.tratt      _Itratt_1-3      (naturally coded; _Itratt_1 omitted)
i.karcateg   _Ikarcateg_1-2   (naturally coded; _Ikarcateg_1 omitted)
i.tumres     _Itumres_1-3     (naturally coded; _Itumres_1 omitted)
i.grf        _Igrf_1-3       (naturally coded; _Igrf_1 omitted)

      failure _d:  status == 1
      analysis time _t:  time

Iteration 0:  log likelihood = -2203.8905
Iteration 1:  log likelihood = -2150.1134
Iteration 2:  log likelihood = -2148.0433
Iteration 3:  log likelihood = -2148.0407
Refining estimates:
Iteration 0:  log likelihood = -2148.0407

Cox regression -- Breslow method for ties

No. of subjects =      476      Number of obs =      476
No. of failures =      403
Time at risk =      561396

Log likelihood = -2148.0407      LR chi2(7) =      111.70
                                Prob > chi2 =      0.0000
```

| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] |
|--------------|------------|-----------|-------|-------|----------------------|
| _Itratt_2 | 1.182177 | .1480009 | 1.34 | 0.181 | .9249487 1.510941 |
| _Itratt_3 | 1.346028 | .1667759 | 2.40 | 0.016 | 1.055816 1.71601 |
| _Ikarcateg_2 | .6716917 | .0908129 | -2.94 | 0.003 | .5153322 .8754929 |
| _Itumres_2 | 2.002355 | .2605405 | 5.34 | 0.000 | 1.55162 2.584026 |
| _Itumres_3 | 2.927033 | .3958184 | 7.94 | 0.000 | 2.245541 3.81535 |
| _Igrf_2 | 1.442294 | .2438163 | 2.17 | 0.030 | 1.035524 2.008849 |
| _Igrf_3 | 1.585478 | .2595517 | 2.82 | 0.005 | 1.150307 2.185278 |

end of do-file

Stata Command

C:\Stata8

E' possibile **confrontare il trattamento CP con il trattamento P** mediante un test opportuno (test di Wald).

Confrontare la sopravvivenza tra gruppi di trattamento differenti da quello di riferimento (CAP):

test _Itratt_2 = _Itratt_3

```
( 1)  _Itratt_2 - _Itratt_3 = 0  
  
      chi2( 1) =      1.13  
      Prob > chi2 =      0.2873
```

Con modelli che contengono un numero elevato di covariate, è preferibile verificare l'assunzione fondamentale di proporzionalità dei rischi mediante **metodi analitici**. E' possibile utilizzare test basati sui **residui di Schoenfeld** dopo avere stimato i parametri del modello.

Verificare l'assunto di proporzionalità dei rischi mediante test basati sui residui di Schoenfeld:

**xi: stcox i.tratt i.karcateg i.tumres i.grf, schoenfeld(sch*) scaledsch(sca*)
stphtest, detail**

Test of proportional hazards assumption

Time: Time

| | rho | chi2 | df | Prob>chi2 |
|--------------|----------|-------|----|-----------|
| _Itratt_2 | 0.00557 | 0.01 | 1 | 0.9100 |
| _Itratt_3 | 0.06599 | 1.81 | 1 | 0.1783 |
| _Ikarcateg_2 | 0.00472 | 0.01 | 1 | 0.9247 |
| _Itumres_2 | -0.08069 | 2.60 | 1 | 0.1070 |
| _Itumres_3 | -0.12097 | 5.61 | 1 | 0.0179 |
| _Igrf_2 | 0.03517 | 0.51 | 1 | 0.4742 |
| _Igrf_3 | 0.08577 | 3.11 | 1 | 0.0778 |
| global test | | 11.27 | 7 | 0.1272 |

Per verificare l'assunzione fondamentale di proporzionalità dei rischi è possibile utilizzare test basati su **covariate tempo-dipendenti** $x_j(t) = x_jg(t)$, dove $g(t)$ può essere la differenza tra il logaritmo della variabile tempo e il logaritmo del tempo medio di sopravvivenza, ovvero $g(t) = [\log(t) - 7]$.

L'obiettivo è quello di riprodurre la tabella 7.3 a pagina 237 del libro di testo.

Verificare l'assunto di proporzionalità dei rischi per la variabile gruppo di terapia mediante un test basato su covariate tempo-dipendenti $TRATT_j * g(TIME)$:

xi: stcox i.tratt i.karcateg i.tumres i.grf, tvc(i.tratt) texp(ln(_t)-7)

est store A

```

do "C:\WINDOWS\TEMP\S
browse
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S
do "C:\WINDOWS\TEMP\S

Stata Results
xi: stcox i.tratt i.karcateg i.tumres i.grf, tvc(i.tratt) texp(ln(_t)-7)
i.tratt      _itratt_1-3      (naturally coded; _itratt_1 omitted)
i.karcateg   _ikarcateg_1-2     (naturally coded; _ikarcateg_1 omitted)
i.tumres     _itumres_1-3      (naturally coded; _itumres_1 omitted)
i.grf        _igrf_1-3        (naturally coded; _igrf_1 omitted)

failure _d: status == 1
analysis time _t: time

Iteration 0:   log likelihood = -2203.8905
Iteration 1:   log likelihood = -2148.3844
Iteration 2:   log likelihood = -2146.1889
Iteration 3:   log likelihood = -2146.1858
Refining estimates:
Iteration 0:   log likelihood = -2146.1858

Cox regression -- Breslow method for ties

No. of subjects =      476          Number of obs =      476
No. of failures =      403
Time at risk   =    561396

Log likelihood = -2146.1858          LR chi2(9) =    115.41
                                   Prob > chi2 =     0.0000

+-----+-----+-----+-----+-----+-----+
|      _t | Haz. Ratio | Std. Err. | z | P>|z| | [95% Conf. Interval] |
+-----+-----+-----+-----+-----+-----+
| rh      |             |            |   |       |                       |
|+-----+-----+-----+-----+-----+-----+
|_itratt_2| 1.198492   | .1805565   | 1.20 | 0.229 | .8920699   | 1.610168 |
|_itratt_3| 1.572847   | .2356725   | 3.02 | 0.003 | 1.172584   | 2.10974  |
|_ikarcateg_2| 6.789344   | .0919272   | -2.86 | 0.004 | .5206858   | .8852785 |
|_itumres_2| 2.021881   | .2637857   | 5.40 | 0.000 | 1.565681   | 2.611007 |
|_itumres_3| 2.950242   | .3997251   | 7.99 | 0.000 | 2.262191   | 3.847565 |
|_igrf_2   | 1.462822   | .2481338   | 2.24 | 0.025 | 1.049071   | 2.039755 |
|_igrf_3   | 1.603931   | .263358    | 2.88 | 0.004 | 1.162579   | 2.212835 |
|+-----+-----+-----+-----+-----+-----+
| t       |             |            |   |       |                       |
|+-----+-----+-----+-----+-----+-----+
|_itratt_2| 1.03002    | .1224529   | 0.25 | 0.804 | .815927    | 1.300289 |
|_itratt_3| 1.248972   | .157441    | 1.76 | 0.078 | .9755588   | 1.599014 |
|+-----+-----+-----+-----+-----+-----+
Note: Second equation contains variables that continuously vary with respect to time; variables are
interacted with current values of ln(_t)-7.

Stata Command

```

xi: stcox i.tratt i.karcateg i.tumres i.grf

lrtest A

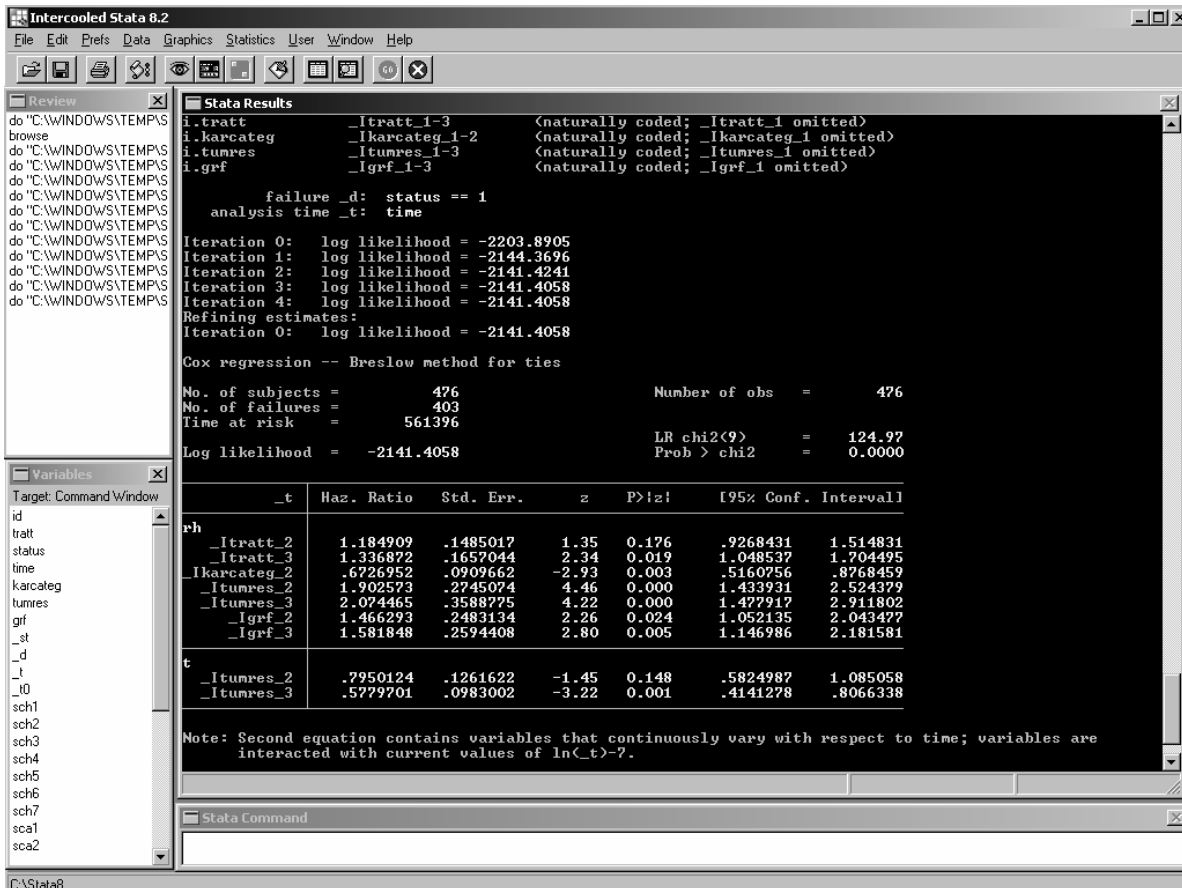
likelihood-ratio test
(Assumption: . nested in A)

LR chi2(2) = 3.71
Prob > chi2 = 0.1565

Verificare l'assunto di proporzionalità dei rischi per la variabile dimensione residuale del tumore mediante un test basato su covariate tempo-dipendenti TUMRES_j*g(TIME):

xi: stcox i.tratt i.karcateg i.tumres i.grf, tvc(i.tumres) texp(ln(_t)-7)

est store B



xi: stcox i.tratt i.karcateg i.tumres i.grf

lrtest B

likelihood-ratio test
(Assumption: . nested in B)

LR chi2(2) = 13.27
Prob > chi2 = 0.0013

Poiché l'assunzione di proporzionalità dei rischi non è verificata per la dimensione residuale del tumore, un modello corretto da adottare è il **modello di Cox stratificato** per questa variabile.

Costruire il modello di Cox stratificato per la dimensione residuale del tumore:

```
xi: stcox i.tratt i.karcateg i.grf, strata(tumres)
```

```
do "C:\WINDOWS\TEMP\S  
browse  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S  
do "C:\WINDOWS\TEMP\S
```

```
Stata Results  
end of do-file  
. do "C:\WINDOWS\TEMP\STD000000000.tmp"  
. xi: stcox i.tratt i.karcateg i.grf, strata(tumres)  
i.tratt       _ltratt_1-3    <naturally coded; _ltratt_1 omitted>  
i.karcateg    _lkarcateg_1-2 <naturally coded; _lkarcateg_1 omitted>  
i.grf         _lgrf_1-3      <naturally coded; _lgrf_1 omitted>  
  
      failure_d: status == 1  
      analysis time _t: time  
  
Iteration 0:   log likelihood = -1747.072  
Iteration 1:   log likelihood = -1736.9286  
Iteration 2:   log likelihood = -1736.7845  
Iteration 3:   log likelihood = -1736.7844  
Refining estimates:  
Iteration 0:   log likelihood = -1736.7844  
  
Stratified Cox regr. -- Breslow method for ties  
No. of subjects =      476      Number of obs =      476  
No. of failures =      403  
Time at risk =      561396  
Log likelihood = -1736.7844  
LR chi2(5) =      20.58  
Prob > chi2 =      0.0010  
  
      _t      Haz. Ratio      Std. Err.      z      P>|z|      [95% Conf. Intervall  
_ltratt_2      1.178878      .1478944      1.31      0.190      .9218973      1.507492  
_ltratt_3      1.317585      .1634501      2.22      0.026      1.033201      1.680245  
_lkarcateg_2    .6735244      .0913078      -2.92      0.004      .5163669      .8785132  
_lgrf_2         1.445924      .2453373      2.17      0.030      1.036854      2.016384  
_lgrf_3         1.566603      .256995      2.74      0.006      1.135854      2.160703  
  
Stratified by tumres  
  
end of do-file
```

```
Stata Command
```

Confrontare la sopravvivenza tra gruppi di trattamento differenti da quello di riferimento (CAP):

```
test _ltratt_2 = _ltratt_3
```

```
( 1) _ltratt_2 - _ltratt_3 = 0
```

```
      chi2( 1) =      0.83  
      Prob > chi2 =      0.3622
```