SPSS

Gold Business Partner

IBM

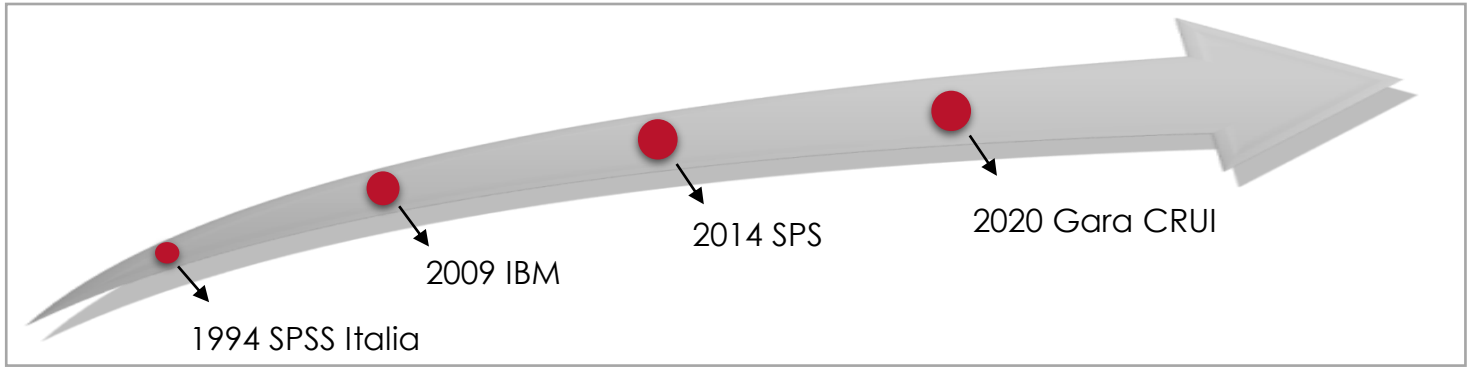
Competency
Data Science & Business Analytics

Webinar

La Biostatistica con IBM SPSS Statistics

Webinar | SPS Srl

Roadmap



Posizionamento

- +25 anni di esperienza sui prodotti SPSS
- IBM BP **Software Support Provider** con competenza IBM Expert in Data Science and Business Analytics
- Gold Business Partner IBM

Presenza sul mercato

- **Scuole e Università**
- **Enti di ricerca (CNR, ISS, CREA ecc.)**
- Sanità (IRCCS, AO, AOU ecc.)
- Pubblica Amministrazione
- Associazioni, Fondazioni
- Aziende (GdO, Finance, Retail, ecc.)

Webinar | CRUI e SPS S.r.l.

Gold Business Partner

Il contratto stipulato tra CRUI e SPS S.r.l. in data 21 ottobre 2019 si riferisce alla fornitura licenze Campus IBM SPSS Statistics del "Catalogo licenze CRUI Campus PA" e di licenze aggiuntive IBM SPSS Amos e IBM SPSS Statistics per usi amministrativi.

Tutte le informazioni sul contratto sono disponibili al link
https://www.spss.it/crui-universita-italiane

Il referente commerciale per l'esecuzione contratto in SPS Srl è:

Laura Zerbini

Responsabile commerciale in Italia per aree Educational e R...

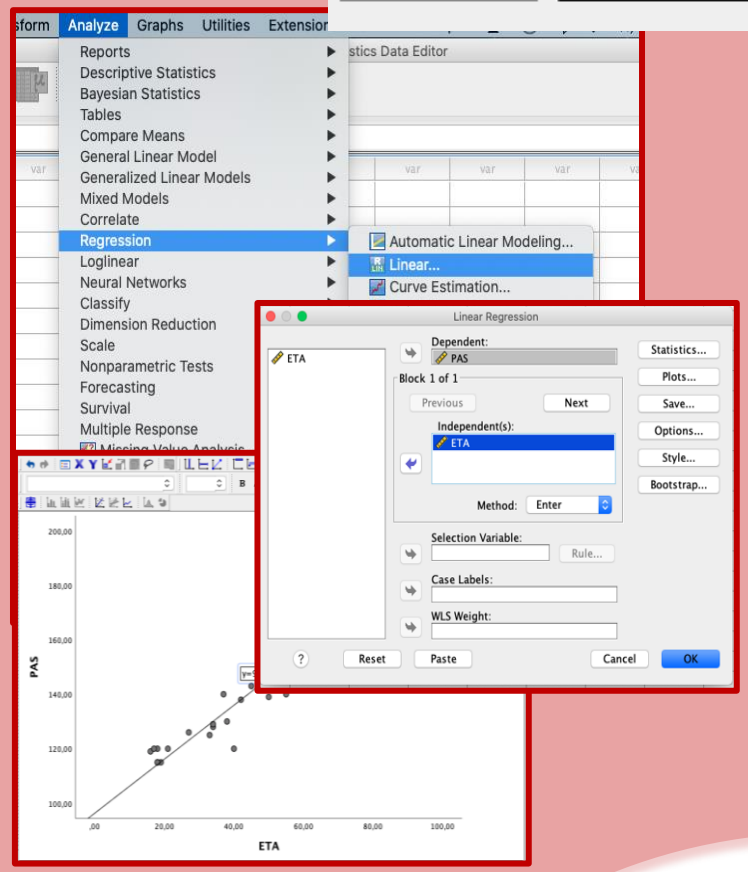
Webinar | IBM SPSS Statistics al lavoro...

Gold Business Partner

Scarica la nostra App sugli store

Download on the App Store

Google play



Antonella Santucci docente @SPS

www.spss.it

Webinar | La Biostatistica con IBM SPSS Statistics

Gold Business Partner

Analisi di sopravvivenza

Principi generali e inserimento dati

Curve di sopravvivenza secondo Kaplan Meier

Calcolare e interpretare le curve di sopravvivenza secondo Kaplan Meier

Confronti fra curve

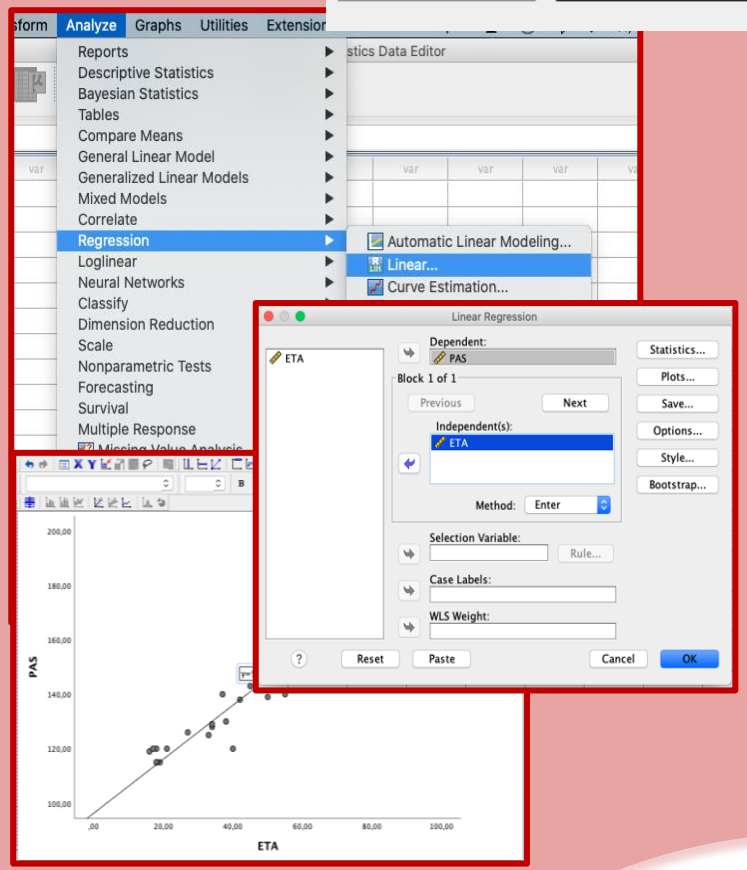
Le statistiche test che consentono confronti fra curve di sopravvivenza

Esempi con IBM SPSS Statistics

Scarica la nostra App sugli store

Download on the App Store

Google play



www.spss.it

- Consiste in tecniche utilizzate quando la variabile oggetto di studio è rappresentata dal tempo intercorrente tra il ***momento iniziale*** definito come origine dell'osservazione e il ***momento finale*** in cui si verifica l'evento (definito come fine della stessa)
- Anche se frequentemente l'evento considerato è il decesso (da cui prende il nome l'insieme delle tecniche), molti altri end points sia sfavorevoli che favorevoli possono essere trattati con le tecniche di analisi della sopravvivenza

2 diverse informazioni sono necessarie:

1) **Se** l'evento si verifica (evento si/no)

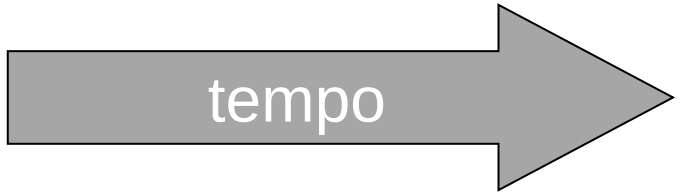
2) **tempo** che intercorre tra data inizio follow-up e data in cui si verifica l'evento (o data dell'ultimo follow-up se l'evento non si verifica)

Variabile di interesse

Tempo che intercorre dall'inizio dell'osservazione fino alla fine del follow up

Inizio del follow-up

(es. Trapianto)



Evento/ Non evento

(es. morte)

Evento =es morte, ripresa del lavoro..

Si assume 1 evento

(nel caso di eventi combinati, quello che arriva per primo)

>1 evento

(es Relapse e Non Relapse Mortality)

incidenza cumulativa a rischi competitivi

Variabile Status

- **Al termine del periodo di osservazione alcuni soggetti hanno registrato l'evento** (decesso / endpoint considerato) **=1**

- **Altri soggetti** o non hanno ancora avuto l'evento o sono usciti dallo studio prima dell'evento, per vari motivi **=0**

i soggetti che, a vario titolo (ancora vivi, withdrawn, lost) non hanno registrato l'evento vengono definiti ***troncati o censored*** (in altre parole incompleti)

La peculiarità di queste tecniche è di gestire i dati *censored*

Censored

- Si assume che l'esperienza dei soggetti persi di vista o usciti vivi alla fine dello studio sia la stessa di coloro che forniscono la risposta durante lo studio
- Ancorché arbitrario, ciò sembra ragionevole per i soggetti usciti vivi, mentre l'esperienza di vita dei soggetti persi di vista potrebbe essere migliore o peggiore di quella dei rispondenti (suscettibile quindi a introdurre bias)

Importante

mantenere al livello più basso possibile i soggetti persi di vista

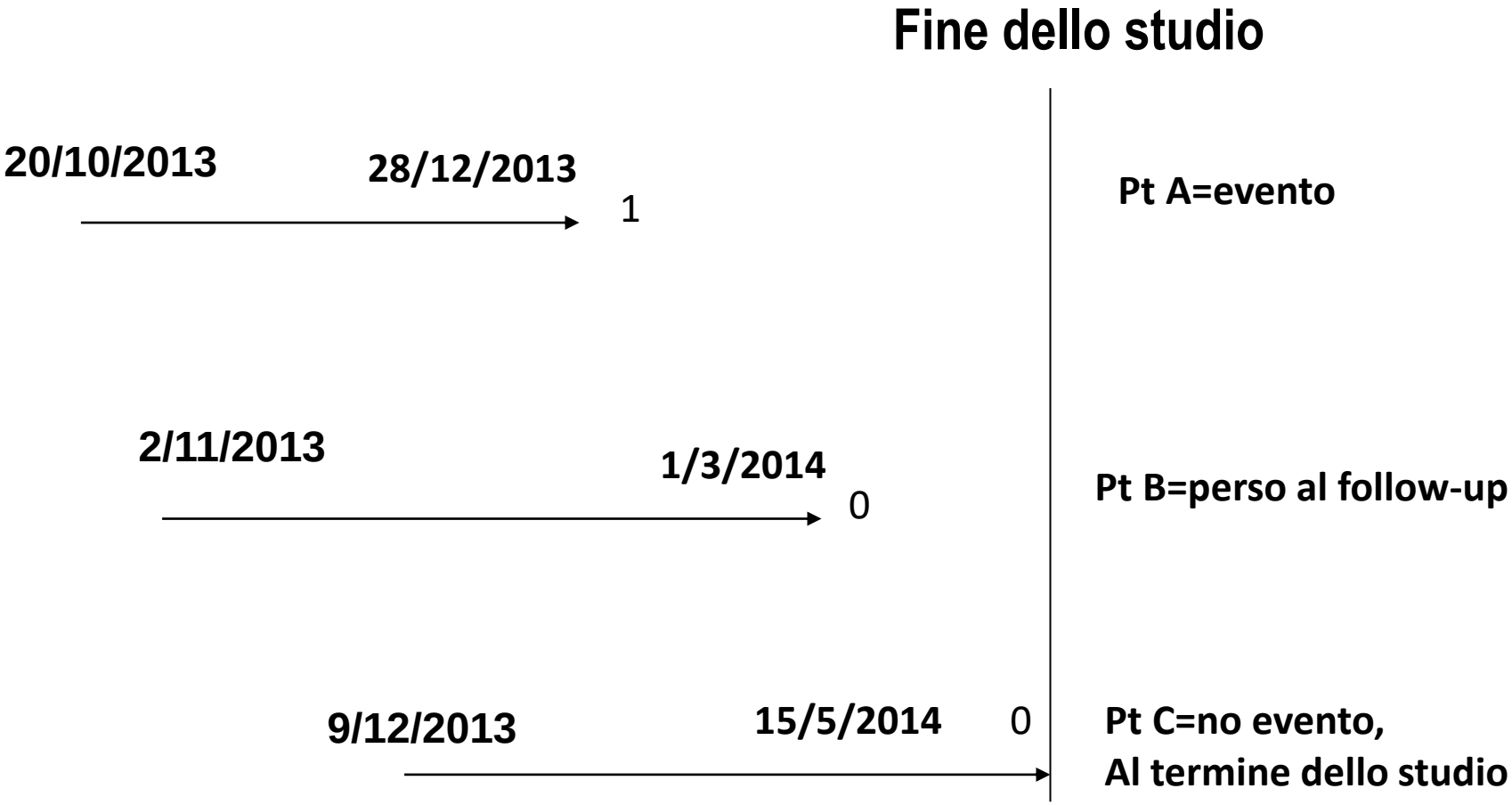
(meno del 5%)

Una buona domanda

Per le osservazioni censored ponetevi questa domanda:

- 'Il fatto che il paziente a un certo punto sia definito censored indica che il suo rischio è diverso rispetto al rischio di un analogo paziente che è ancora nel follow up a quel tempo?'
- Se la risposta è 'no' allora la definizione di censored è corretta.

Le diverse situazioni



Come raccogliere i dati

Tempo di follow-up

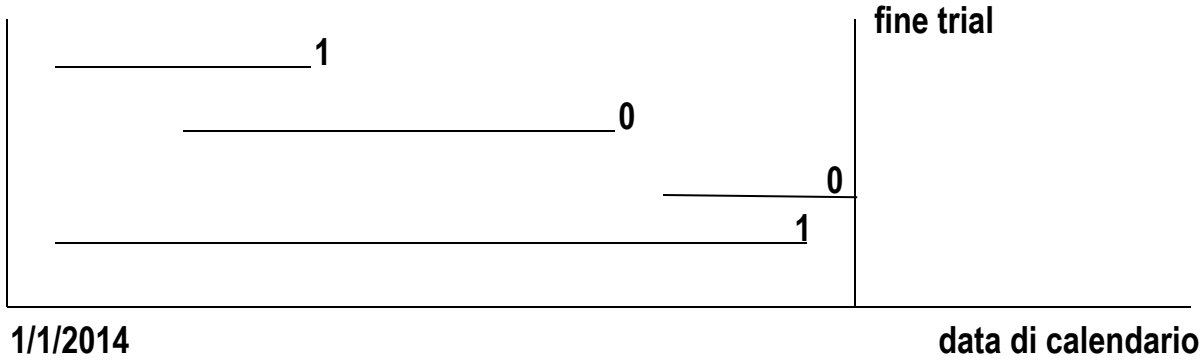
id paziente	Data inizio FU	Data ultimoFU	status
1	20/10/2013	28/12/2013	1(evento)
2	02/11/2013	01/03/2014	0 (perso FU)
3	15/01/2014	18/05/2014	*0(uscito vivo)*
4	22/01/2014	20/02/2014	*1 (evento)*
5	15/03/2014	18/05/2014	*0 (uscito vivo)*

Dati pronti per l'analisi

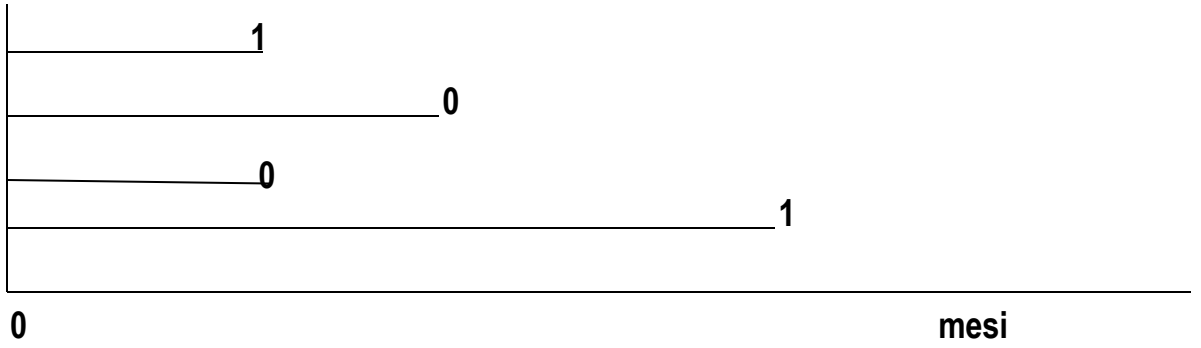
Data ultimo follow-up	Data inizio follow-up	Intervallo di tempo in…	status
28/12/2013	20/10/2013	68	1
01/03/2014	02/11/2013	119	0
15/03/2014	15/01/2014	59	0
20/02/2014	22/01/2014	29	1
28/04/2014	15/04/2014	13	0

Tempo e stato

- Non è necessario che i soggetti entrino in osservazione allo stesso tempo



- Per ciascun soggetto viene infatti calcolato l'intervallo di tempo (in mesi, in giorni, ecc.) in cui è rimasto in osservazione e lo stato in cui si trova al termine

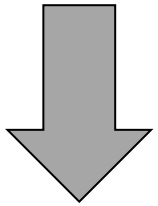


IBM SPSS Statistics

Gold Business Partner

- Aprire file ‘surv.sav’

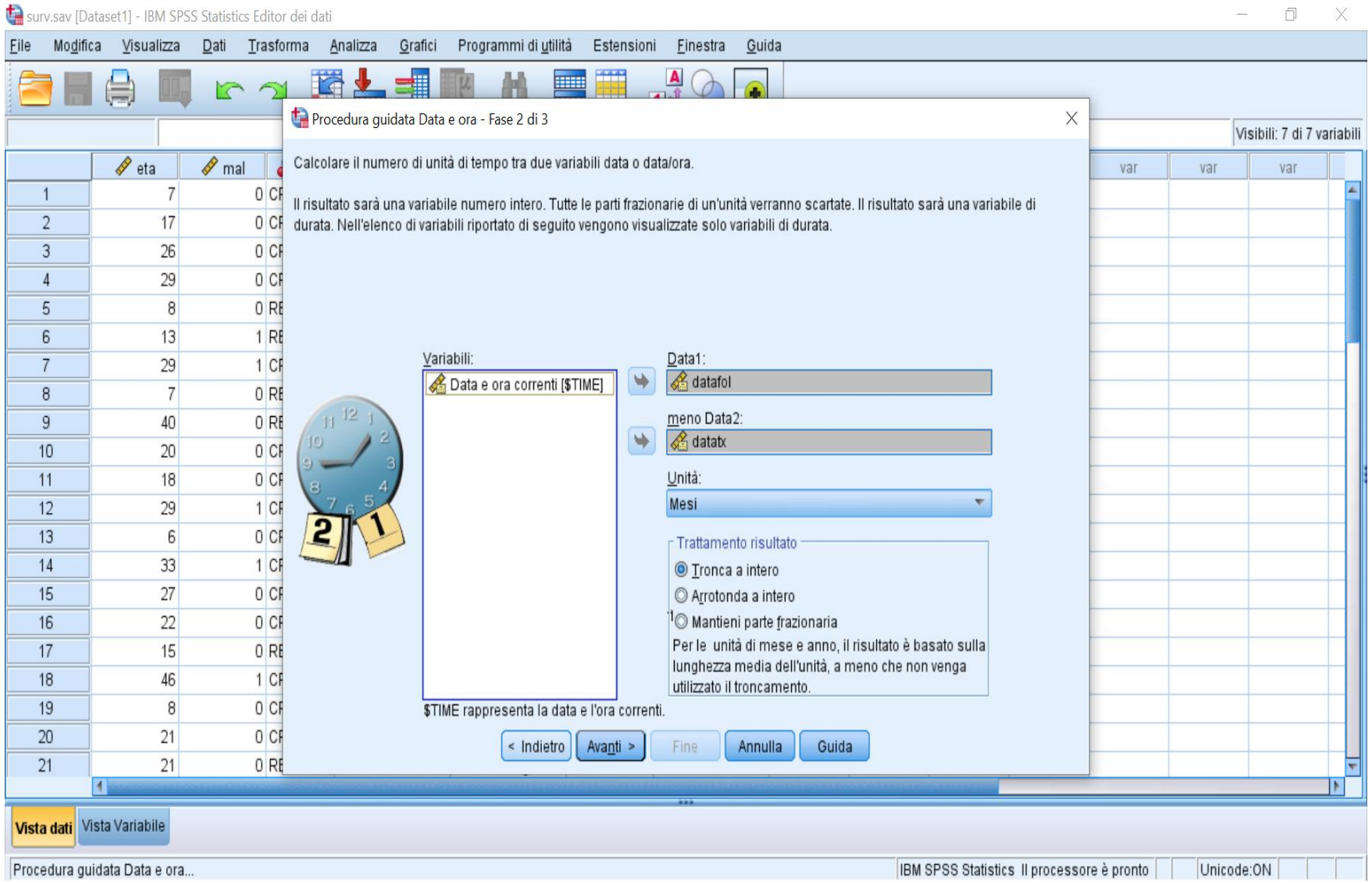
- Trasforma
  procedura guidata data e ora
  eseguire calcoli con date e ore
  calcolare il n. di unita di t tra 2 date



VIENE CREATA LA NUOVA VARIABILE INTERVALLO DI TEMPO

IBM SPSS Statistics

Gold Business Partner



www.spss.it

Analisi della sopravvivenza

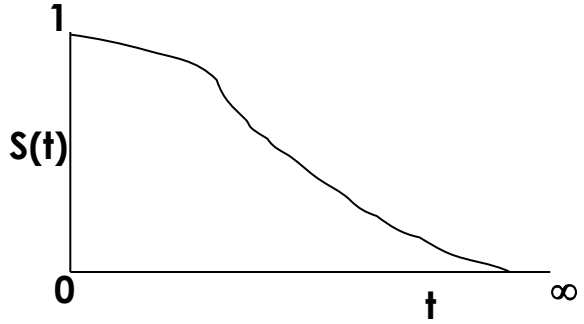
- stimare la curva di sopravvivenza
  - Metodo del prodotto limite
  - Metodo attuariale o life table
- valutare le differenze tra due o più sottogruppi di soggetti
  - Analisi univariata (log rank test)
  - Analisi multivariata (modello a rischi proporzionali di Cox)

La funzione di sopravvivenza S(t)

La funzione di sopravvivenza esprime la probabilità che un individuo sopravviva oltre il tempo “t”, cioè la probabilità che un individuo non abbia ancora subito l’evento critico al tempo t.

Funzione di sopravvivenza

S(t) è la probabilità che una unità sopravviva un tempo > dello specificato tempo t



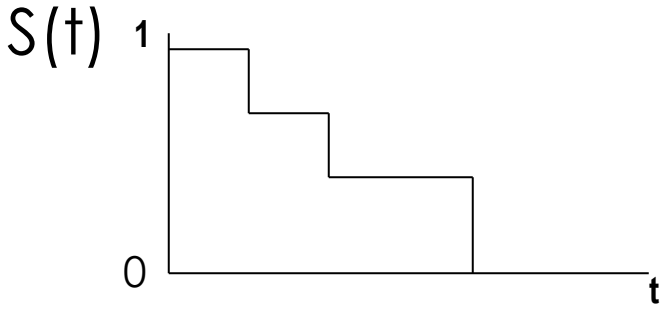
La funzione

- **non è mai crescente**
- **al tempo t = 0 S(t) = S(0) = 1**
- **al tempo t = ∞ S(t) = S(∞) = 0**

Metodo del prodotto limite (Kaplan Meier)

- Stima la curva di sopravvivenza in base al criterio di massima verosimiglianza. La sua peculiarità è di poter utilizzare le informazioni fornite sia dai dati *censored* che dagli eventi

- Si ottiene una funzione a gradini che tiene conto che per arrivare all'ultimo gradino bisogna aver superato con successo i precedenti



- Questo significa stimare le probabilità di sopravvivenza di ciascun gradino e moltiplicarle tra di loro in modo da ottenere la stima cumulativa di tutta la funzione di sopravvivenza

Tempo e stato

- L'informazione tempo in osservazione fornita da ogni soggetto è espressa nell'esempio dalla variabile T , mentre Stato è una variabile binaria riguardante lo stato del paziente: es. assume valore 1 in caso di evento e valore 0 in caso di troncamento.
- I dati ordinati in senso crescente sulla variabile T sono ora pronti per essere analizzati

Caso	T	Stato
1	26	1
3	66	0
2	69	0
4	74	1

Kaplan Meier

Esempio

Pt	Tempo	Status
1	26	1
3	66	1
2	69	0
4	74	1
.	.	.

Al tempo 0 ⟶ S(0) = 1

Al tempo 26 ⟶ S(26) = 1 * (1-1/14) = 0.93

Al tempo 66 ⟶ S(66) = 0.93 * (1-1/13) = 0.86

Al tempo 74 ⟶ S(74) = 086 * (1-1/11) = 0.78

la probabilità S si modifica solo in corrispondenza di eventi

I censored contribuiscono al denominatore finché sono in osservazione

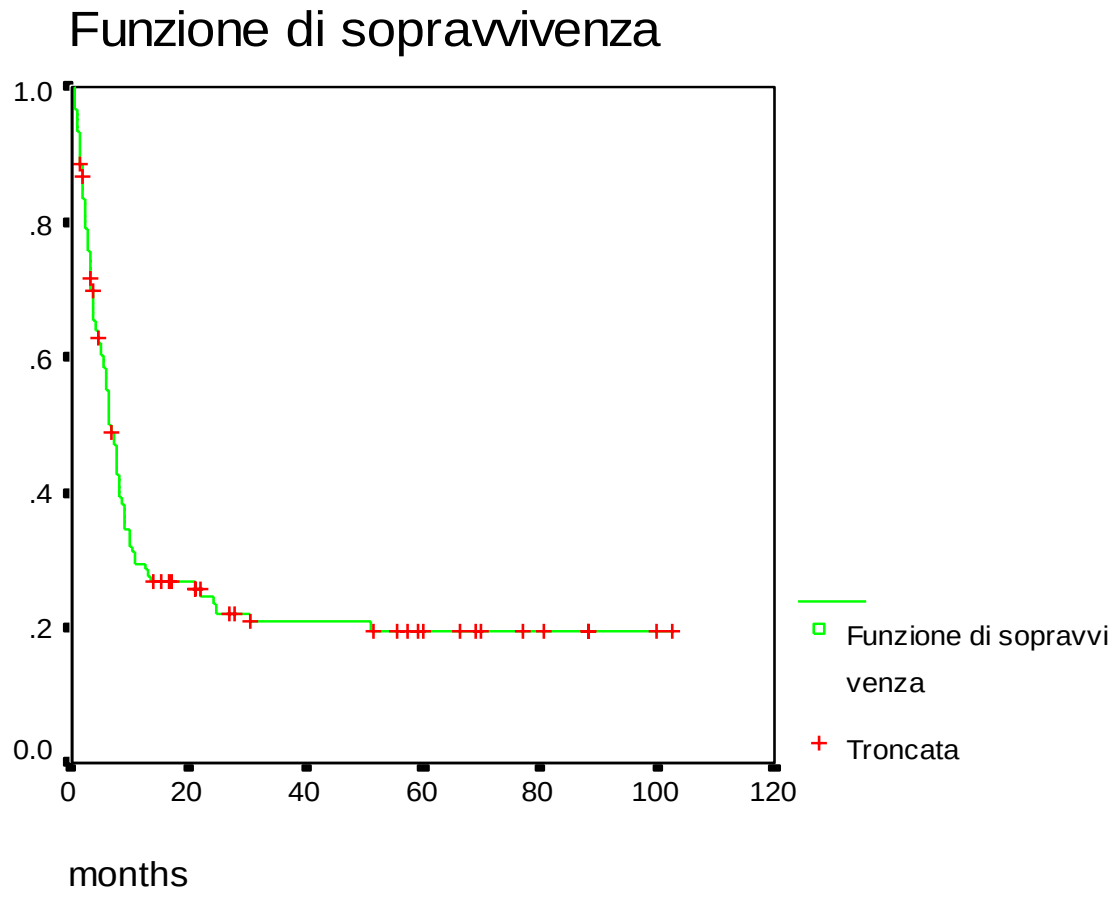
Kaplan Meier

Time (days)	Status	Number at risk	Probability of survival	Standard error
0		14	1.00	
26	1	14	0.93	0.069
66	1	13	0.86	0.094
69*	0	12		
74	1	11	0.78	0.113
82*	0	10		
88	1	9	0.69	0.129
89*	0	8		
117*	0	7		
133*	0	6		
144*	0	5		
172*	0	4		
252*	0	3		
291*	0	2		
305*	0	1		

Gold Business Partner

www.spss.it

Curva di sopravvivenza secondo il metodo di Kaplan Meier

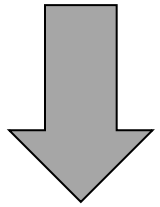


Tempo mediano di sopravvivenza: tempo in cui la probabilità di sopravvivenza assume valore pari a 0,5

Analizza

sopravvivenza

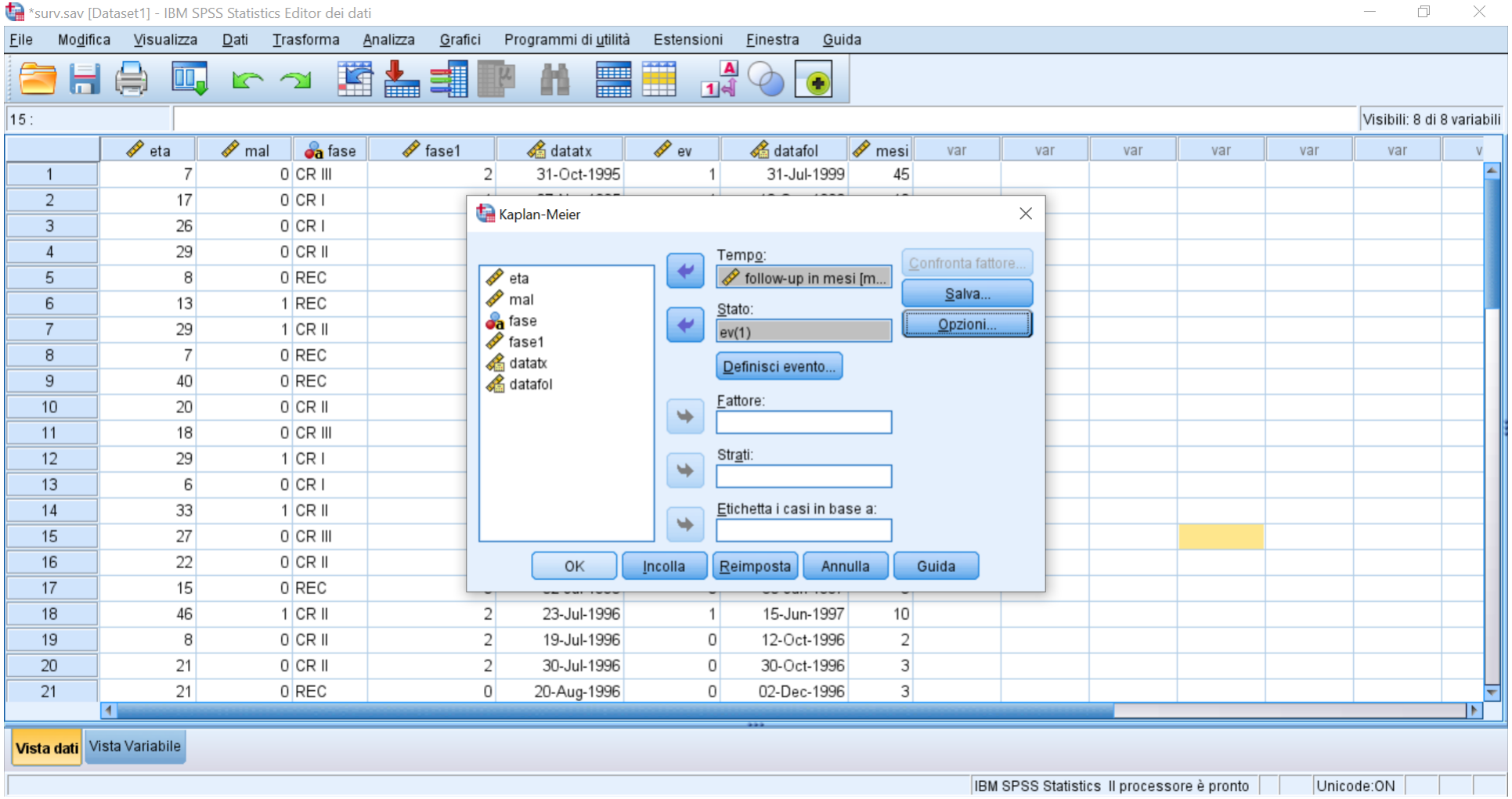
Kaplan Meier



Si va a calcolare la curva di sopravvivenza

IBM SPSS Statistics

Gold Business Partner



	eta	mal	fase	fase1	datatx	ev	datafol	mesi
1	7	0	CR III	2	31-Oct-1995	1	31-Jul-1999	45
2	17	0	CR I					
3	26	0	CR I					
4	29	0	CR II					
5	8	0	REC					
6	13	1	REC					
7	29	1	CR II					
8	7	0	REC					
9	40	0	REC					
10	20	0	CR II					
11	18	0	CR III					
12	29	1	CR I					
13	6	0	CR I					
14	33	1	CR II					
15	27	0	CR III					
16	22	0	CR II					
17	15	0	REC					
18	46	1	CR II	2	23-Jul-1996	1	15-Jun-1997	10
19	8	0	CR II	2	19-Jul-1996	0	12-Oct-1996	2
20	21	0	CR II	2	30-Jul-1996	0	30-Oct-1996	3
21	21	0	REC	0	20-Aug-1996	0	02-Dec-1996	3

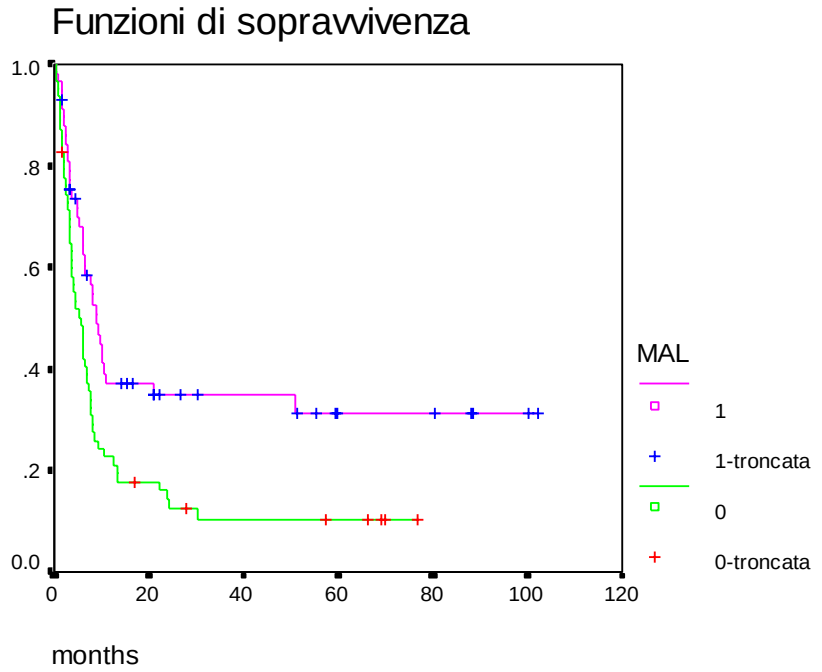
www.spss.it

Differenza statistica tra 2 o più curve di sopravvivenza

- Per verificare l'ipotesi nulla "non ci sono differenze di distribuzione del rischio di eventi all'interno di una variabile con 2 o piu' livelli" viene utilizzato un test non parametrico:

Log-Rank test

- E' un test non parametrico in quanto non viene assunto alcun tipo di distribuzione della funzione di rischio *h(t)*



Long Rank Test

Confronto tra due o più curve di sopravvivenza

- Si supponga di avere N_A pazienti al trattamento A e N_B pazienti al trattamento B. In un dato giorno sono sotto osservazione n pazienti di cui 3/5 appartenenti al gruppo A e 2/5 al gruppo B. Sotto l'ipotesi nulla di uguale efficacia dei trattamenti ci si attende che dei **d** eventi capitati in quel giorno i 3/5 appartengano al gruppo A e i 2/5 al gruppo B. Questo ripetuto per tutti i tempi in cui si verificano eventi
- Se ci si discosta in modo significativo dal risultato atteso significa che A e B hanno effetti diversi sulla sopravvivenza (cioè non sono ininfluenti)
- L'ipotesi nulla viene testata attraverso

$$\chi^2_{1g.l.} = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_B - E_B)^2}{E_B}$$

O= osservati
E=attesi

Altri test statistici

- **WILCOXON** test (detto **BRESLOW test in IBM SPSS Statistics**) mette più enfasi nella parte iniziale delle curve, dove il numero di soggetti a rischio è maggiore. Viene dato più peso agli eventi che si verificano all'inizio del follow-up

- **TARONE WARE** idem

- .....

Criticità circa KM

- KM stima la probabilità di sopravvivenza nei diversi tempi ma
- NON stima **l'entità** della differenza di surv nei # sottogruppi
- NON stima l'effetto di variabili **quantitative**
- NON stima **l'effetto simultaneo di più variabili**

⇒ **ANALISI MULTIVARIATA**

Webinar | Importazione e preparazione dei dati

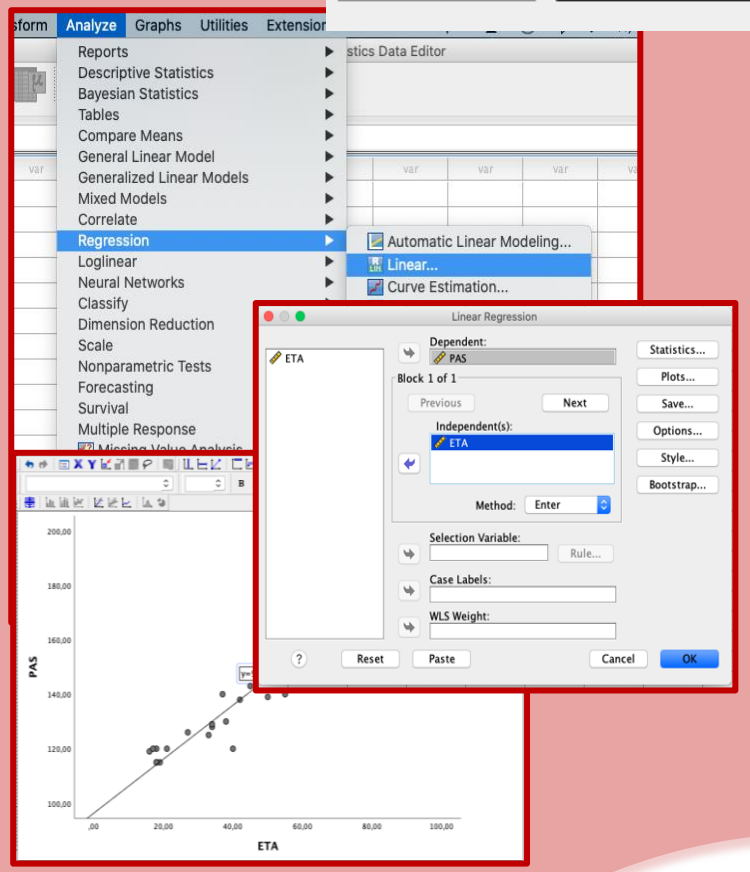
Gold Business Partner

Scarica la nostra App sugli store

Download on the App Store

Google play

Ringraziamenti e Conclusioni



LAURA ZERBINI @SPS

www.spss.it