

### 3.8.1 Analisi delle variabili associate con il metodo di Kaplan-Meier

Per ogni caso selezionato nei dati dell'IPIMC abbiamo la durata in giorni trascorsa dalla nascita alla morte (dato completo) o alla data di uscita dallo studio (dato troncato) e possiamo analizzare i tempi di sopravvivenza della SD in funzione di alcune covariate di interesse.

Attraverso il metodo non parametrico di Kaplan-Meier possiamo descrivere le caratteristiche del processo in esame senza fare ipotesi circa la distribuzione degli eventi e tenendo conto dei tempi di sopravvivenza troncati (censored) in cui non si è verificato l'evento.

Il numero e la dimensione degli intervalli non sono stabiliti esogenamente dal ricercatore, come invece avviene nella stima attuariale, ma derivano dall'ordinamento stesso dei tempi di accadimento. Il metodo di calcolo delle tavole di sopravvivenza con cui viene eseguita la stima delle probabilità di sopravvivenza e delle relative curve di sopravvivenza avviene in ogni punto in cui si verifica un evento e non in base alla suddivisione del periodo di osservazione in intervalli di tempo uguali e predefiniti.

In ciascun punto temporale in cui si è verificato un evento vengono stimate le probabilità condizionate e attraverso il limite di tali probabilità si arriva alla stima del tasso di sopravvivenza in ciascun punto (per questo motivo tale metodo è anche denominato product-limit); attraverso il modello di Kaplan-Meier possiamo confrontare anche il tasso di sopravvivenza globale in gruppi diversi e possiamo tracciare le funzioni di sopravvivenza o di rischio confrontando graficamente i risultati ottenuti in modo da ricavare informazioni più dettagliate circa l'andamento della sopravvivenza della SD.

Questo metodo non consente di stimare direttamente il rischio; si possono ottenere le stime della funzione di rischio tramite l'applicazione di tecniche di liscio (smoothing) sul rischio cumulato.

La variabile  $T$  che misura il tempo di sopravvivenza è una variabile casuale continua, e i vari tempi  $t_1, t_2, \dots, t_p$  sono i tempi in cui si osserva un decesso in uno dei soggetti del collettivo; le variabili secondo cui si vogliono analizzare le differenze di sopravvivenza possono essere categoriali o continue.

La tavola di sopravvivenza stima il tempo, la sopravvivenza cumulata, l'errore standard, gli eventi cumulati e rimanenti, il tempo di sopravvivenza media (nel nostro caso non possiamo calcolare la sopravvivenza mediana) l'errore standard e l'intervallo di confidenza del 95%.

Per verificare l'uguaglianza o meno delle curve di sopravvivenza relative a diverse categorie del fattore preso in esame vengono utilizzati due test non parametrici: il Log-rank test e il test di Breslow; il Log-rank test accentua le differenze nelle funzioni di sopravvivenza alla fine del processo in esame e il test di Breslow accentua le differenze all'inizio. Entrambi i test si distribuiscono secondo un  $\chi^2$  con  $m-1$  gradi di libertà (dove  $m$  è il numero di gruppi messi a confronto). L'ipotesi nulla  $H_0$  è che le funzioni di sopravvivenza siano uguali per tutte le categorie dello stesso gruppo ovvero che non vi siano differenze di sopravvivenza tra i gruppi considerati.

Si possono calcolare i grafici delle funzioni di sopravvivenza, delle funzioni di rischio e del logaritmo delle funzioni di sopravvivenza.

I risultati delle elaborazioni proposte di seguito sono stati ottenuti tramite il programma statistico SPSS (versione 10.0).

Abbiamo considerato le covariate relative al neonato e ai genitori al momento della nascita del figlio disponibili nei dati nell'IPIMC che potessero rappresentare possibili fattori esplicativi della sopravvivenza delle persone con la SD.

La Tabella 3.21 riporta le principali variabili che sono state analizzate mostrando, per ogni variabile, il numero di casi validi su cui si basa l'analisi, il numero di decessi associati, la percentuale di dati troncati e il valore della significatività del Log-rank test con il metodo di Kaplan-Meier.

**Tabella 3.21 - Variabili associate alla sopravvivenza della SD con il metodo di Kaplan-Meier**

Covariate	n	Decessi	% censored	p (Log-rank test)
Sesso	1996	438	78,06	0,8332
Zona di residenza	1988	430	78,37	0,0000
Coorte di nascita	1996	438	78,06	0,0011
Malformazioni associate	1996	438	78,06	0,0000
Prematuro	1881	411	78,15	0,0003
Sottopeso	1951	418	78,58	0,0001
Piccolo per Età Gestazionale	1845	397	78,48	0,0001
Farmaci in gravidanza	1018	198	80,55	0,4537
Malattie in gravidanza	1018	198	80,55	0,6173
Fumo in gravidanza	1018	198	80,55	0,6600
Gravidanza gemellare	1996	438	78,06	0,1461
Età della madre	1992	437	78,06	0,0511
Età del padre	1888	411	78,23	0,3466
Titolo di studio della madre	1746	373	78,64	0,9176
Titolo di studio del padre	1734	373	78,49	0,0854

Tra le caratteristiche del neonato abbiamo considerato il sesso (maschi e femmine), la zona di residenza (Nord, Centro e Sud), la coorte di nascita (1978- 1983, 1984- 1989 e 1990- 1994), la presenza di malformazioni associate (nessuna, lieve, grave e cardiaca), la prematurità (nato prima della 37<sup>a</sup> settimana e nato a termine), il basso peso alla nascita (inferiore o superiore a 2.500 grammi), l'essere o meno PEG (vedi paragrafo 3.5.4).

Tra le caratteristiche relative alla gravidanza abbiamo considerato l'uso di farmaci in gravidanza (si, no), la presenza di malattie durante la gravidanza (si, no), l'aver fumato durante la gravidanza (si, no) e il genere del parto (singolo, doppio).

Tra le caratteristiche dei genitori al momento della nascita del figlio abbiamo considerato l'età ( $\leq 24$ , 25- 34, 35- 44, 45- 54,  $\geq 55$ ) e il titolo di studio (diploma di scuola elementare, diploma di scuola media, diploma di scuola superiore e laurea).

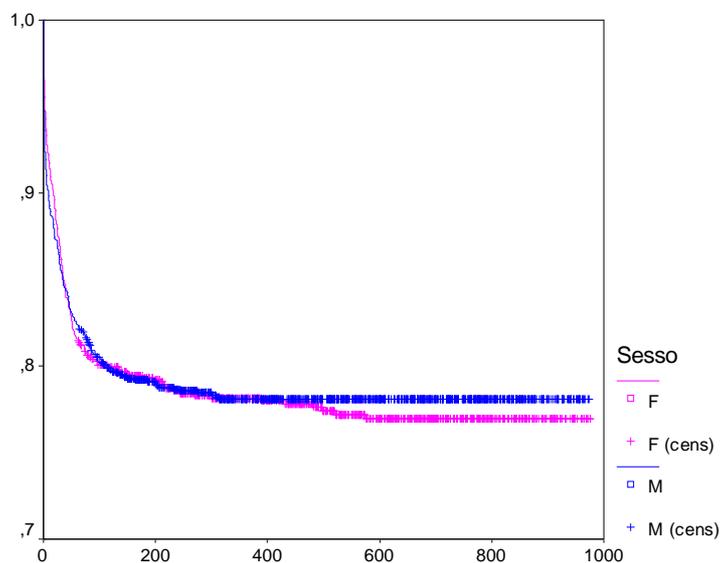
Sebbene le differenze di genere rispetto alla sopravvivenza della SD non siano significative secondo i dati dell'IPIMC, analizzando le curve di sopravvivenza<sup>1</sup> riportate nel Grafico 3.2 possiamo notare che i maschi

<sup>1</sup> In tutti i grafici relativi alle curve di sopravvivenza l'asse dei tempi è espresso in settimane

hanno una mortalità più bassa rispetto alle femmine; tale risultato è spesso emerso anche in letteratura (vedi paragrafo 3.1).

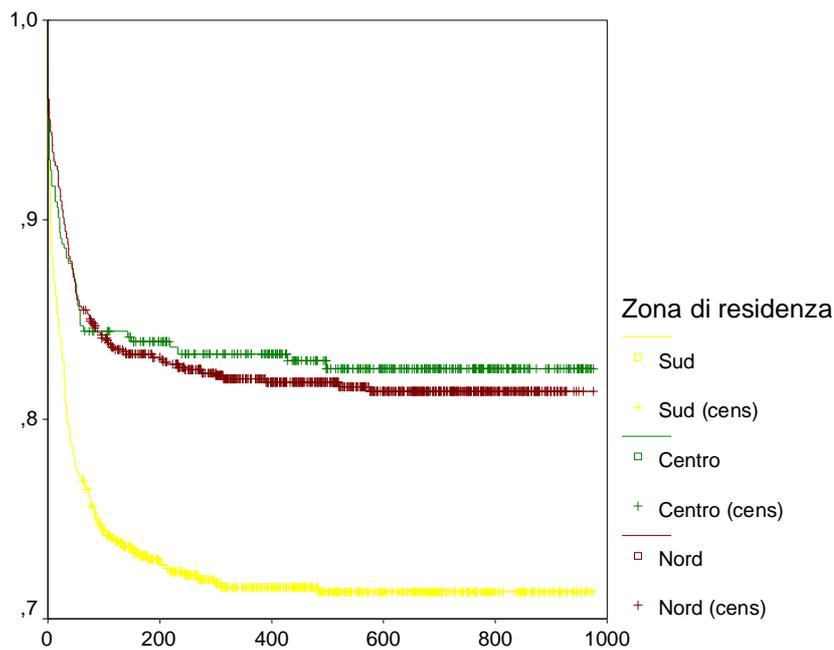
Molto significativa risulta, invece, la zona di residenza del neonato; il Grafico 3.3 mostra le curve di sopravvivenza relative a questa variabile. La sopravvivenza del Nord e del Centro è molto simile (con il Centro che mostra una mortalità leggermente inferiore) mentre il Sud presenta una mortalità delle persone con SD molto più alta. Tale risultato era già emerso dallo studio dei dati dell'IPIMC relativi al periodo 1978- 1984 (Mastroiacovo – 1992).

**Grafico 3.2 – Curva di sopravvivenza per la variabile sesso**



La forbice significativa nella sopravvivenza delle persone con SD residenti nel Nord e nel Centro rispetto a quelle che risiedono al Sud è quasi sicuramente da attribuirsi alle diverse risorse presenti sul territorio, alla diversa qualità dell'assistenza medica e riabilitativa, ma anche alla diminuita possibilità di accesso ai servizi da parte delle fasce della popolazione più disagiate (Mastroiacovo – 1992).

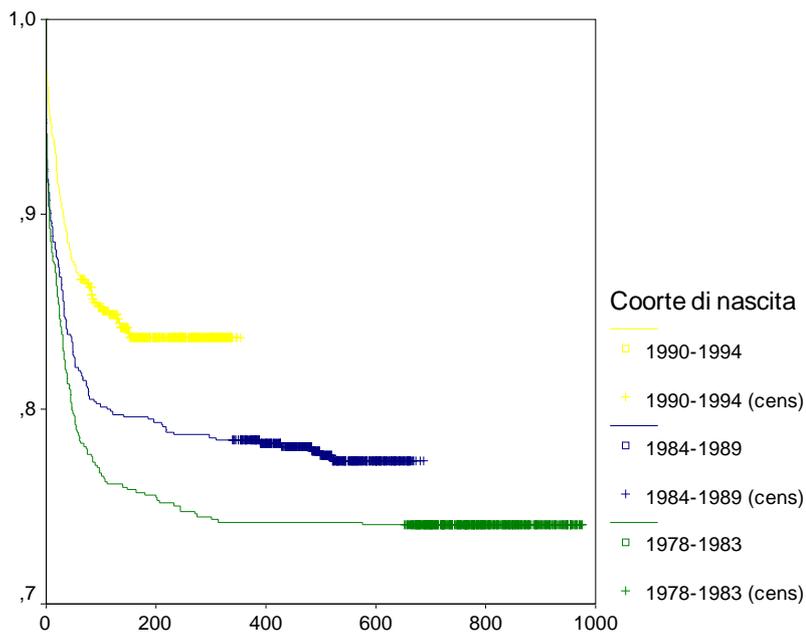
**Grafico 3.3 – Curva di sopravvivenza per la variabile Zona di residenza**



Anche le differenze nella sopravvivenza delle persone con la SD tra le diverse coorti di nascita risultano significative; il Grafico 3.4 mostra come la mortalità diminuisca notevolmente passando dalla coorte di nascita del 1978- 1983 alla coorte di nascita del 1990- 1994.

Ovviamente i miglioramenti in campo clinico che sono stati sperimentati nel corso del tempo hanno portato benefici in termini di sopravvivenza sia per la popolazione totale sia per la popolazione delle persone con la SD. Le coorti di nascita più “giovani vengono osservate per un periodo minore rispetto alle più “anziane”.

**Grafico 3.4 – Curva di sopravvivenza rispetto alla variabile Coorte di nascita**



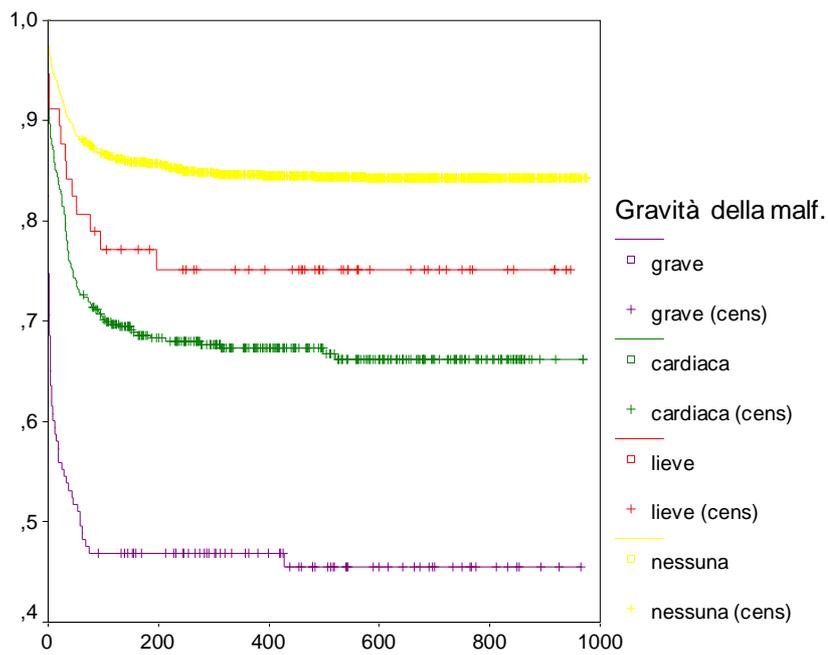
La presenza di malformazioni associate alla SD comporta notevoli differenze nella sopravvivenza futura delle persone con la SD. Dal Grafico 3.5 si nota che all'aumentare della gravità della malformazione associata alla SD aumenta la mortalità e le differenze che si riscontrano sono significative.

Nonostante i miglioramenti nel trattamento delle malformazioni cardiache abbiano diminuito la mortalità infantile, queste rimangono comunque un importante fattore predittivo della mortalità (vedi paragrafo 3.2).

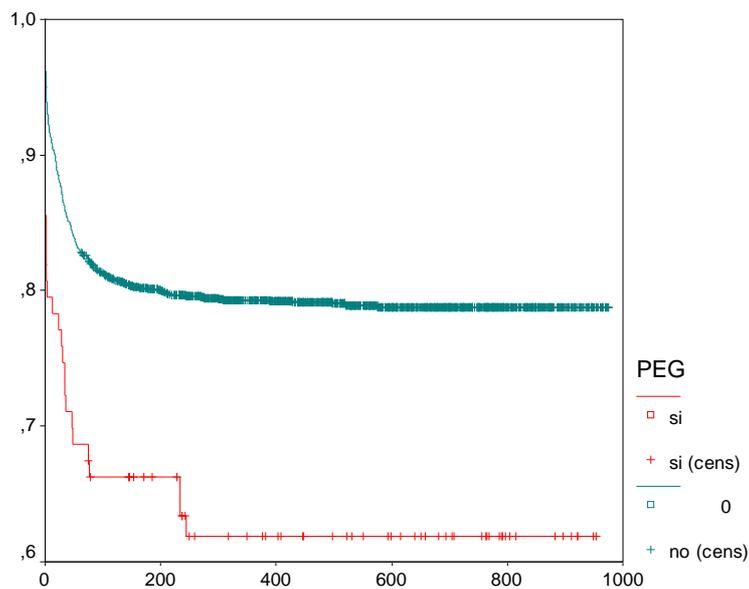
Sia l'essere nati prematuri sia l'essere nati sottopeso (peso alla nascita inferiore a 2.500 grammi) sono variabili significative rispetto alla sopravvivenza della SD. Anche la variabile che racchiude in una sola caratteristica le due variabili appena citate, ovvero l'essere Piccoli per Età Gestazionale (PEG), risulta essere un importante fattore predittivo della mortalità relativa alla SD.

Il Grafico 3.6 mostra le due curve di sopravvivenza relative ai bambini che nascono Piccoli per Età Gestazionale e ai bambini che nascono con un peso adeguato all'epoca gestazionale indipendentemente dal fatto di essere nati a termine o meno. La differenza nelle due curve, oltre ad essere significativa, è anche molto accentuata.

**Grafico 3.5 – Curva di sopravvivenza rispetto alla variabile Malformazioni associate alla SD**



**Gráfico 3.6 – Curva di sopravvivenza rispetto alla variabile Piccoli per Età Gestazionale**



Rispetto alla sopravvivenza della SD abbiamo che lo stimatore di Kaplan-Meier non mostra differenze significative rispetto a nessuna delle caratteristiche della gravidanza analizzate (farmaci, malattie, fumo e gravidanze gemellari).

Neanche l'età e l'istruzione dei genitori al momento della nascita del figlio ha implicazioni significative sulla sopravvivenza delle persone con la SD.

Dopo aver valutato separatamente l'effetto delle singole covariate sulla probabilità di sopravvivenza della SD attraverso i dati dell'IPIMC, ci proponiamo ora di analizzare l'effetto congiunto dei diversi fattori predittivi della mortalità della SD tenendo conto della potenziale influenza degli altri.

Poiché vi è difficoltà nel formulare ipotesi esaustive e specifiche per individuare un modello parametrico che rappresenti la distribuzione dei tempi di sopravvivenza relativi alla SD, abbiamo scelto di utilizzare il modello semiparametrico di Cox.

### 3.8.2 Il modello semiparametrico di Cox

Il fenomeno che ci interessa studiare, ovvero la sopravvivenza della SD, può assumere solo due stati: vivo/morto e nei dati dell'IPIMC abbiamo a disposizione solo variabili che ipotizziamo time-constant, rilevate al momento della nascita del bambino con la SD.

Nei modelli semiparametrici non viene fatta nessuna ipotesi sulla funzione di sopravvivenza, e quindi sul rischio, mentre viene riparametrizzata l'influenza delle covariate sul rischio; viene specificata solo la forma funzionale per valutare l'influenza delle covariate sui tempi di sopravvivenza e la forma dell'andamento del rischio viene lasciata non specificata.

La funzione di rischio (che rappresenta la probabilità di morire al tempo  $t$  per un individuo con vettore di covariate  $z$  sopravvissuto fino al tempo  $t$ ) viene generalmente fattorizzata in due parti:

$$\lambda(t/z) = \lambda_0(t)H(z,\beta) \quad (3.7)$$

dove  $\lambda_0(t)$  è la funzione di rischio di base per un individuo con vettore di covariate  $z=0$  che dipende solo da  $t$  ed è lasciata non specificata, e  $H(z,\beta)$  dipende solo dalla matrice delle covariate che individuano le possibili influenze sul rischio  $\lambda_0(t)$  e  $\beta$  sono i coefficienti associati alle covariate (non è presente l'intercetta poiché è "assorbita" dal rischio di base). L'effetto delle covariate può solo indurre traslazioni proporzionali sul rischio ma non può modificarne l'andamento; siamo interessati a studiare se e quanto i valori delle variabili esplicative diversi da zero modificano la funzione di sopravvivenza, ipotizzando un effetto moltiplicativo.

La forma scelta da Cox per esplicitare l'effetto delle covariate sul rischio di base è quella esponenziale in cui  $H(z,\beta) = \exp(\beta z)$ .

Attraverso la funzione lineare nei parametri  $\beta$  su scala esponenziale si esplicita l'espressione regressiva molto semplice tra rischio e covariate; tale modello mette in relazione il logaritmo del rischio con le variabili esplicative:

$$\log \frac{\lambda(t,z)}{\lambda_0(t)} = \beta z = \beta_1 z_1 + \beta_2 z_2 + \dots + \beta_p z_p \quad (3.8)$$

Il parametro  $e^{\beta_i}$  misura la variazione del rischio in relazione alla differenza tra la modalità in esame e la modalità di riferimento a parità delle altre covariate. I parametri  $\beta$  indicano l'incremento o la riduzione percentuale del rischio su scala logaritmica dovuto alla modalità di una variabile relativamente al rischio di base che si è scelto per la modalità di riferimento in cui  $\exp(\beta z) = 1$  se  $z=0$ , a parità delle altre covariate. Se le covariate non hanno alcun effetto sul rischio, se cioè  $\beta_1 = \beta_2 = \dots = \beta_p = 0$ , avremo  $\lambda(t/z) = \lambda_0(t)$ .

Il modello di regressione di Cox è un metodo che consente di gestire i casi in cui non si è verificato l'evento e di ottenere stime dei coefficienti per ciascuna delle covariate inserite nel modello; è certamente il modello statistico più utilizzato in ambito biomedico.

Tale modello può essere applicato solo se è verificata l'ipotesi di proporzionalità dei rischi nei sottinsiemi di dati. Il rapporto dei rischi in diversi gruppi deve essere proporzionale nei diversi tempi di sopravvivenza ovvero non deve variare nel tempo (per questo motivo tale modello viene anche detto *dei rischi proporzionali*).

Quando si scelgono le variabili esplicative da inserire nel modello, possono essere incluse solo le variabili che presentano un rapporto tra le funzioni di rischio relative alle diverse modalità costanti nel tempo; le curve devono mantenere una distanza tra loro costante nel tempo.

Per verificare graficamente che vi sia proporzionalità nei rischi relativi alle diverse categorie di una variabile si può considerare la funzione  $\log[-\log S_i(t)]$ , dove  $S_i(t)$  è la funzione di sopravvivenza relativa alla categoria  $i$ ; se è vera l'ipotesi di proporzionalità le curve delle diverse modalità devono essere parallele nel tempo o quantomeno non devono mai incrociarsi.

L'ipotesi della proporzionalità dei rischi, pur consentendo notevoli semplificazioni per ciò che concerne gli aspetti inferenziali di stima e verifica di ipotesi sui parametri del modello di Cox, risulta a volte in contrasto con l'andamento reale del fenomeno. Non è raro infatti che l'effetto dovuto ad un particolare fattore di rischio vari nel tempo; tale ipotesi va comunque sempre verificata prima di applicare il modello di Cox.

La stima dei parametri incogniti  $\beta$  del modello e le relative procedure di verifica di ipotesi avvengono tramite il metodo della massima verosimiglianza parziale (per la fattorizzazione di Cox) che consente di introdurre esplicitamente il troncamento.

La funzione di verosimiglianza per il parametro  $\beta$  è data da:

$$L(\beta) = \prod_{i \in D} \frac{\lambda_i(t_{(k)})}{\sum_{j \in R_k} \lambda_j(t_{(k)})} = \prod_{i \in D} \frac{\exp(\beta' z_i)}{\sum_{j \in R_k} \exp(\beta' z_j)} \quad (3.9)$$

In questa espressione  $t_{(k)}$  indica il k-imo tempo di fallimento osservato,  $i$  identifica l'individuo che ha sperimentato l'evento in  $t_{(k)}$ ,  $R_k$  è l'insieme degli individui ancora a rischio un istante prima di  $t_{(k)}$  e  $D$  è l'insieme degli individui che hanno sperimentato l'evento durante il periodo di osservazione (osservazioni complete non censurate). Questa funzione, sebbene non possa essere considerata una funzione di verosimiglianza associata ad un reale esperimento statistico, mantiene tutte le proprietà che caratterizzano una funzione di verosimiglianza ordinaria.

I tempi troncati non danno un contributo diretto alla verosimiglianza, essendo la produttoria limitata agli effettivi tempi di fallimento osservati; ciascun termine della produttoria può essere interpretato come la probabilità che, osservato un fallimento in  $t_{(k)}$ , sia stato proprio l' $i$ -imo individuo tra coloro ancora esposti al rischio a sperimentare l'evento.

Grazie alle proprietà asintotiche del metodo della massima verosimiglianza è possibile costruire test per verificare l'ipotesi nulla che  $\beta_i=0$ .

La variabile che misura il tempo di sopravvivenza deve essere continua; le variabili esplicative possono essere categoriali o continue (se sono categoriali devono essere codificate in forma di indicatori ovvero rappresentate da una o più variabili dicotomiche di comodo).

Le covariate inserite nel modello di Cox sono le stesse relative alle caratteristiche del neonato analizzate tramite il metodo di Kaplan-Meier descritto nel paragrafo precedente (avevamo sottolineato che nessuna delle variabili relative alle caratteristiche della gravidanza e dei genitori risultavano associate alla sopravvivenza della SD).

Anche in questo caso è stato utilizzato il pacchetto statistico SPSS (versione 10.0).

Prima di iniziare l'analisi statistica dei risultati del modello di Cox abbiamo verificato graficamente per tutte le covariate scelte che vi fosse proporzionalità tra i rischi per tutta la durata del follow-up. I grafici della funzione  $\log[-\log S_i(t)]$  ci suggeriscono che i rischi sono proporzionali e quindi è possibile applicare il modello di Cox per tutte le variabili considerate.

Il primo passo nella descrizione del modello di Cox applicato è quello di assumere un soggetto di riferimento (osservato o teorico) caratterizzato dal fatto di avere tutti i valori delle covariate identicamente nulli; tramite il modello di Cox non viene stimato il rischio assoluto di una modalità ma il rischio relativo tra le modalità considerate e la modalità di riferimento. Per semplicità di lettura abbiamo scelto l'individuo che si suppone abbia il rischio di morte più basso, ricodificando opportunamente i relativi valori delle variabili esplicative.

Nella Tabella 3.22 sono riportate le modalità delle variabili inserite nel modello di Cox (per ogni variabile è indicata la modalità di riferimento) con il relativo valore del parametro  $\beta$ , la sua significatività, il valore del parametro  $e^\beta$  che ci consente una lettura più immediata dei risultati e l'intervallo di confidenza di  $e^\beta$  al 95%.

**Tabella 3.22 - Risultati del modello di Cox sui dati dell'IPIMC**

Covariate	$\beta$	p	exp( $\beta$ )	IC 95%
Sesso: femmina (rif. maschio)	0,105	0,355	1,111	0,889-1,387
Zona di residenza: Centro (rif. Nord)	-0,110	0,527	0,896	0,637-1,259
Zona di residenza: Sud (rif. Nord)	0,716	0,000	2,047	1,598-2,622
Coorte di nascita: 1978-83 (rif. 1990-1994)	0,766	0,000	2,152	1,591-2,912
Coorte di nascita: 1984-89 (rif. 1990-1994)	0,411	0,007	1,508	1,116-2,037
Gravità malf. associata: lieve (rif. nessuna)	0,460	0,159	1,584	0,835-3,008
Gravità malf. associata: cardiaca (rif. nessuna)	1,097	0,000	2,994	2,319-3,866
Gravità malf. associata: grave (rif. nessuna)	1,686	0,000	5,399	3,873-7,526
Prematuro: si (rif. no)	0,242	0,130	1,274	0,931-1,744
Peso inferiore a 1.500 gr: si (rif. no)	0,654	0,109	1,924	0,865-4,277
PEG: si (rif. no)	0,470	0,059	1,599	1,001-2,555

Il sesso non risulta significativo anche se i maschi mostrano una sopravvivenza leggermente migliore rispetto alle femmine.

La sopravvivenza di coloro che risiedono al Centro non è significativamente diversa da coloro che vivono al Nord (anche se risulta leggermente superiore). Per coloro che risiedono al Sud rispetto a coloro che risiedono al Nord, invece, vi è una sensibile differenza in termini di sopravvivenza della SD e, infatti, la mortalità dei residenti al Sud risulta due volte maggiore rispetto ai residenti al Nord.

Questo conferma quanto già emerso nelle analisi precedenti: le condizioni di sopravvivenza sono simili tra Nord e Centro e sono nettamente migliori rispetto al Sud.

La mortalità relativa alla prima coorte di nascita disponibile nei dati dell'IPIMC (1978- 1983) è 2 volte superiore rispetto a quella relativa all'ultima coorte (1990- 1994); la coorte intermedia (1984- 1989) presenta una mortalità 1,5 volte superiore rispetto all'ultima. Tutte queste differenze sono ampiamente significative.

L'aver una malformazione associata alla SD è un importante fattore predittivo della mortalità. Mentre non vi sono differenze significative rispetto alla sopravvivenza della SD se non si ha nessuna malformazione associata alla SD o se la malformazione associata è lieve, quando si ha anche una malformazione cardiaca o grave il rischio aumenta notevolmente e le differenze di sopravvivenza sono molto significative. In particolare, quando si ha una malformazione cardiaca associata alla SD il rischio di morte è 3 volte superiore rispetto a quando non vi è nessuna malformazione; quando la malformazione associata è grave il rischio è più che quintuplicato.

Ciò fa riflettere sul fatto che, sebbene i progressi nella medicina generale e nella cardiocirurgia in particolare abbiano contribuito notevolmente a migliorare la sopravvivenza della popolazione totale e della popolazione con la SD, le malformazioni congenite rappresentano ancora un importante fattore predittivo della mortalità della SD.

Non risultano significative le variabili relative alle condizioni cliniche del neonato al momento della nascita: l'essere prematuri, sottopeso o Piccoli per Età Gestazionale.

Cercheremo ora di approfondire l'analisi sulla sopravvivenza della SD relativa al primo mese e al primo anno di vita attraverso l'uso dei modelli logistici.

### 3.8.3 Il modello logistico

Dopo aver considerato i possibili fattori predittivi della mortalità della SD in tutto il periodo di osservazione del follow-up dell'IPIMC, siamo ora interessati ad approfondire il ruolo degli stessi fattori in due segmenti particolarmente delicati: il primo mese e il primo anno di vita.

Nei dati dell'IPIMC tutti i casi selezionati sono stati osservati per almeno un anno. Il numero di decessi di bambini con SD che si verifica nel primo mese è di 134 su 438 totali (30,6%); i decessi di bambini con SD che si verificano nel primo anno di vita sono 339 su 438 (77,4%).

Poiché il numero di decessi che si verifica nel primo mese, nonostante rappresenti una quota ampia rispetto al totale dei decessi in tutto il periodo di osservazione, è limitato e poiché l'analisi dei tempi di accadimento degli eventi viene condotta in giorni dobbiamo gestire un numero considerevole di *ties* (eventi che accadono lo stesso giorno), abbiamo preferito approfondire la mortalità della SD nel primo mese di vita attraverso l'uso di modelli a tempo discreto.

Quando si utilizzano modelli a tempo continuo, come ad esempio il modello di Cox, la probabilità che si verifichino *ties* è zero, quindi la loro presenza può portare a notevoli distorsioni nelle stime. Per poter confrontare i risultati che si ottengono quando si considera il primo anno di vita, abbiamo applicato le stesse metodologie anche il questo periodo del ciclo di vita.

I modelli a tempo discreto presuppongono che l'evento di interesse accada in punti discreti del tempo; poiché la nostra variabile di risposta è binaria (vivo/morto) abbiamo applicato la regressione logistica. I modelli logistici a tempo discreto possono essere applicati ai dati utilizzando la regressione logistica dei pacchetti statistici comuni dopo aver modificato opportunamente il set di dati nell'archivio "persona- periodo".

Quando utilizziamo modelli a tempo discreto abbiamo che la funzione di rischio è rappresentata dalla probabilità condizionata di subire l'evento di interesse al tempo  $t_i$ , dato che l'evento non si è ancora verificato al tempo  $t_{i-1}$ :

$$\lambda(t_i) = \Pr(T = t_i / T \geq t_i) \quad (3.10)$$

Nel modello logistico vengono modellizzati gli Odds (rapporto di probabilità) delle probabilità condizionate. Il modello considera che, per ogni individuo appartenente alla popolazione in esame, gli Odds del verificarsi dell'evento al tempo  $t$  (con  $t=1,2,\dots$ ) sono proporzionali agli Odds di un individuo di base, ovvero:

$$\frac{\lambda(t/z_i)}{1-\lambda(t/z_i)} = \frac{\lambda_0(t)}{1-\lambda_0(t)} \exp\left[\sum_{j=1}^p \beta_j z_{ijt}\right] \quad (3.11)$$

dove la funzione del rischio di base è data da:

$$\lambda_0(t) = \frac{1}{1 + \exp(-\alpha_t)} \quad (3.12)$$

I parametri del modello sono dati dai coefficienti di regressione  $\beta$  delle covariate  $z$  e dai coefficienti  $\alpha$  del rischio di base. Per l'individuo di base tutte le covariate assumono il valore 0.

I modelli logistici sono modelli non parametrici poiché non presuppongono nessun legame funzionale tra il logaritmo dell'Odds e i parametri. Se l'Odds di una covariata è uguale a 1 vuol dire che la covariata non è influente rispetto all'accadimento dell'evento di interesse.

Il rapporto tra Odds (Odds Ratio) è dato da:

$$\psi = \frac{\lambda(t/1)/[1-\lambda(t/1)]}{\lambda(t/0)/[1-\lambda(t/0)]} = \exp(\beta) \quad (3.13)$$

da cui si ricava che il logaritmo dell'Odds Ratio è il parametro  $\beta$  che misura l'associazione tra il rischio e la covariata  $z$ .

Vediamo ora i risultati relativi all'analisi condotta sul primo mese e sul primo anno di vita della mortalità della SD secondo i dati dell'IPIMC.

### 3.8.3.1 *Il primo mese di vita*

Abbiamo considerato le stesse variabili utilizzate nel modello di Cox; l'archivio "persona- periodo" relativo alla sopravvivenza nel primo mese di vita passa da 1.996 individui con SD a 7.655 persone- anno.

La Tabella 3.23 riporta per ogni modalità delle variabili inserite nel modello (tra parentesi si trova la modalità scelta come riferimento) il valore del parametro  $\beta$ , la sua significatività, il valore del parametro  $e^\beta$  che ci consente una lettura più immediata dei risultati e l'intervallo di confidenza di  $e^\beta$  al 95%.

**Tabella 3.23 - Risultati del modello logistico nel primo mese di vita con i dati dell'IPIMC**

Covariate	$\beta$	p	exp( $\beta$ )	IC 95%
Sesso: femmina (rif. maschio)	-0,283	0,146	0,753	0,514-1,104
Zona di residenza: Centro (rif. Nord)	0,410	0,125	1,506	0,893-2,542
Zona di residenza: Sud (rif. Nord)	1,023	0,000	2,781	1,801-4,293
Coorte di nascita: 1978-83 (rif. 1990-1994)	1,554	0,000	4,729	2,666-8,389
Coorte di nascita: 1984-89 (rif. 1990-1994)	1,128	0,000	3,091	1,765-5,410
Gravità malf. associata: lieve (rif. nessuna)	1,017	0,035	2,765	1,072-7,136
Gravità malf. associata: cardiaca (rif. nessuna)	1,418	0,000	4,129	2,601-6,555
Gravità malf. associata: grave (rif. nessuna)	2,585	0,000	13,006	8,112-20,851
Prematuro: si (rif. no)	1,005	0,000	2,731	1,768-4,219
Peso inferiore a 1.500 gr: si (rif. no)	1,696	0,000	5,453	2,600-11,438
PEG: si (rif. no)	1,164	0,000	3,202	1,723-5,950

Il sesso non è significativamente influente neanche rispetto alla mortalità della SD nel primo mese di vita.

I risultati rispetto alla zona di residenza concordano con i risultati del modello di Cox relativo a tutto il periodo di osservazione; il Nord e il Centro hanno una sopravvivenza della SD simile (la differenza non è significativa) mentre il Sud mostra valori della mortalità significativamente maggiori rispetto al Nord (il rischio al Sud è quasi il triplo).

Anche la coorte di nascita mostra un andamento del rischio simile a quello che assumeva nel modello di Cox. La sopravvivenza migliora notevolmente nel tempo essendo quasi cinque volte superiore nella prima coorte di nascita disponibile (1978- 1983) rispetto all'ultima (1990- 1994); la coorte intermedia (1984- 1989) mostra una sopravvivenza tre volte superiore rispetto sempre all'ultima coorte considerata.

La presenza di una malformazione associata alla SD comporta differenze significative nei rischi di mortalità della SD durante il primo mese di vita qualsiasi sia la gravità. Se la malformazione associata è lieve il rischio aumenta di 2,7 volte; se c'è una malformazione cardiaca la mortalità è quattro volte superiore rispetto ai neonati che non presentano nessuna malformazione. Se la malformazione associata alla SD è grave il rischio di morte nel primo mese di vita è addirittura tredici volte superiore.

Tutte le variabili che si riferiscono alle condizioni cliniche del neonato al momento della nascita comportano differenze significative nella sopravvivenza della SD nel primo mese di vita.

Se si è prematuri il rischio di morte nel primo mese è 2,7 volte superiore rispetto a se si è nati a termine. Quando il peso alla nascita è inferiore a 1.500 grammi la mortalità è 5,4 volte più alta. Se si nasce Piccoli per Età Gestazionale si rischia 3,2 volte di più di morire entro il primo mese di vita.

Vediamo ora i risultati dello stesso modello in cui vengono considerate le stesse variabili ma il periodo di riferimento è il primo anno di vita.

### 3.8.3.2 *Il primo anno di vita*

L'archivio "persona- periodo" relativo al primo anno di vita conta 20.850 persone- anno. La Tabella 3.24 riassume i risultati dell'applicazione

del modello logistico descritto nel paragrafo 3.8.3 avendo inserito nel modello le stesse variabili che erano state considerate per il modello di Cox.

**Tabella 3.24 - Risultati del modello logistico nel primo anno di vita con i dati dell'IPIMC**

Covariate	$\beta$	p	exp( $\beta$ )	IC 95%
Sesso: femmina (rif. maschio)	0,299	0,053	1,349	1,009-1,803
Zona di residenza: Centro (rif. Nord)	-0,289	0,226	0,749	0,469-1,196
Zona di residenza: Sud (rif. Nord)	0,713	0,000	2,040	1,486-2,802
Coorte di nascita: 1978-83 (rif. 1990-1994)	0,741	0,000	2,099	1,435-3,069
Coorte di nascita: 1984-89 (rif. 1990-1994)	0,306	0,119	1,538	0,925-1,995
Gravità malf. associata: lieve (rif. nessuna)	0,252	0,585	1,287	0,521-3,179
Gravità malf. associata: cardiaca (rif. nessuna)	1,065	0,000	2,900	2,092-4,020
Gravità malf. associata: grave (rif. nessuna)	1,407	0,000	4,083	2,522-6,612
Prematuro: si (rif. no)	0,183	0,391	1,201	0,790-1,826
Peso inferiore a 1.500 gr: si (rif. no)	0,106	0,887	1,111	0,259-4,767
PEG: si (rif. no)	0,221	0,524	1,247	0,633-2,456

Il sesso continua a non avere influenza sulla sopravvivenza della SD anche se si considera il primo anno di vita.

La zona di residenza mantiene i rischi simili nel Nord e nel Centro e segnala, ancora una volta, livelli peggiori al Sud (doppi rispetto al Nord) anche se leggermente più bassi se confrontati con i livelli nel primo mese di vita in cui il rischio del Sud era quasi il triplo di quello del Nord.

Anche la coorte di nascita presenta lo stesso andamento del rischio nel primo anno di vita rispetto a quello nel primo mese di vita anche se le differenze sono minori; mentre nel primo mese di vita il rischio nella prima coorte di nascita era quasi cinque volte superiore rispetto al rischio relativo all'ultima coorte di nascita, quando si analizza la mortalità nel primo anno di vita si rileva un rischio relativo alla prima coorte analizzata che risulta solo due volte superiore rispetto all'ultima coorte.

Ciò può essere dovuto ai maggiori progressi che si sono registrati recentemente nelle terapie rivolte alle cure perinatali rispetto ai progressi nelle cure a vantaggio del primo anno di vita. Non si registrano differenze significative nella sopravvivenza della SD nel primo anno di vita tra la coorte dei nati nel 1984- 1989 e la coorte di nascita 1990- 1994.

Anche per quanto riguarda la presenza di malformazioni associate alla SD, quando si analizza la sopravvivenza nel primo anno di vita, non vi sono differenze significative tra i bambini che non hanno nessuna malformazione associata e tra coloro che hanno una malformazione lieve. Permangono invece importanti differenze quando la malformazione è cardiaca o grave e i rischi di morte della SD nel primo anno di vita sono rispettivamente tre e quattro volte superiori rispetto a quando non c'è nessuna malformazione.

Quando si considera complessivamente la sopravvivenza della SD nel primo anno di vita non risultano più significative come fattori predittivi le variabili che indicano le condizioni cliniche alla nascita ovvero l'essere prematuri, sottopeso o Piccoli per Età Gestazionale. Ciò è probabilmente dovuto al fatto che queste variabili sono determinanti appena nati mentre perdono di importanza quando si cresce; tali variabili non erano significative neanche nel modello di Cox che considerava tutto il periodo di osservazione del follow-up.

Le minori associazioni tra la sopravvivenza della SD e le variabili osservate che vengono riscontrate nel primo anno di vita rispetto a quelle che sono state riscontrate nel primo mese di vita, possono essere dovute anche al fatto che quanto più si aggregano i dati tanto più si appiattiscono le differenze tra le variabili.