Ministero dello Sviluppo Economico

DIPARTIMENTO PER LO SVILUPPO E LA COESIONE ECONOMICA

NUCLEO TECNICO DI VALUTAZIONE E VERIFICA DEGLI INVESTIMENTI PUBBLICI

UVER – UNITÀ DI VERIFICA

VISTO

Visualizzazione Interattiva della Stima dei Tempi delle Opere

Nota metodologica

I modelli di durata applicati alle fasi attuative di un'opera pubblica

Novembre, 2010

Numero identificativo documento/versione:	1603-01
Data di aggiornamento:	4 Novembre 2010

Indice

1 INTRODUZIONE

In questa nota verranno illustrati i principali aspetti metodologici dei modelli statistici che sono alla base dello strumento VISTO. Questa particolare famiglia di modelli, generalmente indicata **come modelli di durata** o **modelli di sopravvivenza**, ha permesso di sfruttare al massimo l'informazione disponibile nelle banche dati sui tempi di attuazione delle singole fasi, poiché consentono di analizzare dati relativi sia a fasi già completate sia a fasi ancora in corso.

2 I MODELLI DI DURATA PER L'ANALISI DI DATI DI SOPRAVVIVENZA

I modelli di durata, o modelli di sopravvivenza, vengono utilizzati nell'analisi di dati che si riferiscono al tempo intercorso tra un istante iniziale e l'accadimento di un evento di interesse.

I più comuni esempi di applicazioni di questo tipo di modelli si trovano in ambito **epidemiologico** (in cui l'evento iniziale è solitamente la diagnosi di una certa patologia o l'inizio della somministrazione di una terapia e l'evento di interesse è la guarigione), oppure in **studi demografici** (per esempio nell'analisi della fecondità, dove l'evento iniziale è il matrimonio o comunque l'inizio dell'unione, oppure la nascita di un figlio, e l'evento di interesse è la nascita di un (altro) figlio, oppure negli studi di mortalità).

Tuttavia questi modelli trovano applicazione anche in altri ambiti, come per esempio quello **economico,** e più in generale nell'analisi di dati di tipo longitudinale.

Nel lavoro qui presentato i modelli di durata sono stati utilizzati per analizzare i tempi di realizzazione delle fasi procedurali di un progetto di investimento.

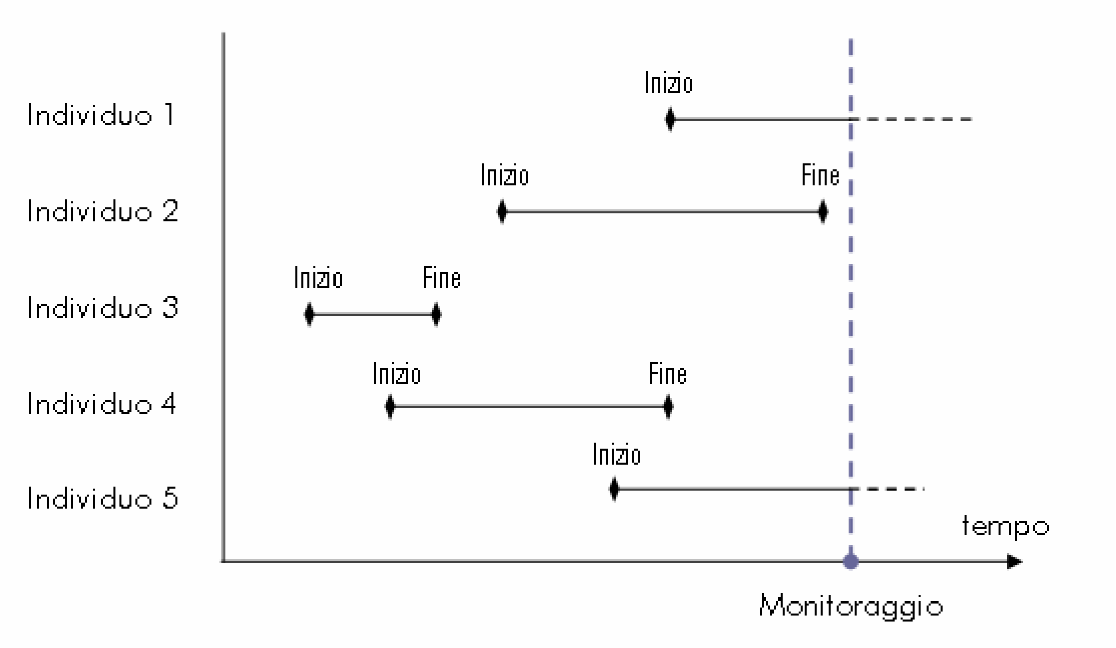
Sono stati stimati 5 modelli distinti, uno per ogni fase procedurale di interesse (Progettazione Preliminare, Definitiva ed Esecutiva, Aggiudicazione del Bando, Esecuzione dei Lavori): in ogni modello, la variabile considerata è la durata della fase calcolata a partire da un evento iniziale (**l'inizio della fase**) fino all'evento conclusivo di interesse, che è rappresentato **dall'inizio della fase successiva**.

Nelle durate sono stati quindi inclusi anche **i tempi di attraversamento** tra una fase e la successiva[1] (interfasi), ad eccezione però della fase dei **Lavori**, per la quale si è considerata la durata tra inizio e fine della fase.

[1] Cfr. I tempi di attuazione delle opere pubbliche

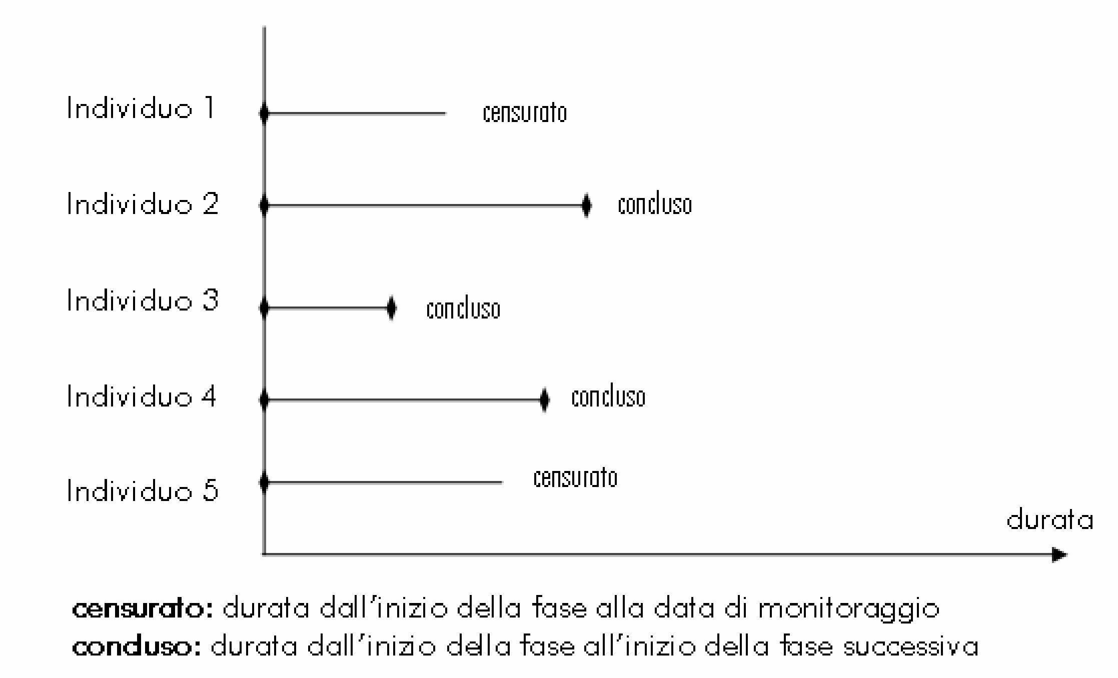
- Per questo tipo di dati i metodi tradizionali di analisi (modelli di regressione lineare) risultano inappropriati, in particolare per due motivi:generalmente, i dati di durata hanno una distribuzione non simmetrica (tendono ad avere una maggior frequenza di durata più brevi e lunghe code, in corrispondenza delle durate più lunghe); per questo l'ipotesi di normalità della distribuzione, alla base dei modelli di regressione lineare, non è applicabile;
- per alcuni individui, le durate osservate possono essere incomplete (**censurate**): è possibile infatti che al momento della rilevazione l'evento oggetto di studio non sia ancora stato sperimentato; i modelli di durata consentono di sfruttare l'informazione, sia pur parziale, anche per questa categoria di individui (cfr. Collett, 1994).

Figura 1 - Esempi di durate concluse e censurate



Le durate possono essere rappresentate su una dimensione temporale 'relativa': per ogni *individuo* (nel nostro caso progetti di investimento) si registra il tempo osservato fino all'evento di interesse, se si è verificato, oppure se non si è verificato, fino all'ultimo istante di osservazione

Figura 2 - Esempi di durate concluse e censurate: durate relative



I modelli di durata permettono di studiare la dinamica temporale con cui gli individui oggetto di studio (nel nostro caso Progetti di investimento) sperimentano l'evento di interesse.

3 ALCUNE DEFINIZIONI: LE FUNZIONI CARATTERISTICHE

Due funzioni sono di particolare interesse nell'analisi della sopravvivenza: la funzione di sopravvivenza $S(t)$ e la funzione di rischio $h(t)$.

Possiamo considerate la durata $\boldsymbol{t}$ come il valore di una variabile casuale $\boldsymbol{T}$ che misura il tempo di attesa fino al verificarsi dell'evento di interesse.

1) Si definisce *funzione di sopravvivenza* la probabilità che un individuo sperimenti l'evento di interessere oltre un certo tempo $\boldsymbol{t}$; formalmente:

$$S(t) = \Pr(T \geq t) \tag{1}$$

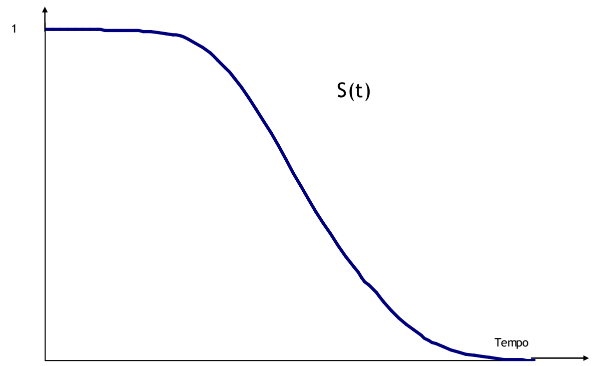
Se indichiamo con $\boldsymbol{f(t)}$ la funzione di densità della variabile casuale $\boldsymbol{T}$, la funzione di sopravvivenza può essere anche espressa come:

$$S(t) = 1 - \int_0^t f(u)du = 1 - F(t) \qquad (2)$$

dove

$$F(t) = \int_0^t f(u)du = \Pr(T < t) \qquad (3)$$

Figura 3 - Funzione di sopravvivenza



Il grafico descrive come nel tempo un insieme di individui (esposti al rischio di sperimentare un certo evento) gradualmente si riduca man mano che nel tempo alcuni individui sperimentano l'evento di interesse.

2) La <u>funzione di rischio</u> ***h(t)*** misura la probabilità che un individuo, che non lo abbia sperimentato prima di quell'istante, sperimenti l'evento di interesse al tempo t; ***h(t)*** si può interpretare come la probabilità istantanea che si verifichi un certo evento che ancora non si è verificato.

In formule

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{\Pr\{t \le T < t + \delta t \mid T \ge t\}}{\delta t} \right\} \qquad (4)$$

La (4) può essere riscritta come:

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{1}{\delta t} \left(\frac{\Pr(t \le T < t + \delta t)}{\Pr(T \ge t)} \right) \right\}$$

che corrisponde a:

$$h(t) = \lim_{\delta t \to 0} \left\{ \frac{F(t + \delta t) - F(t)}{\delta t} \right\} \frac{1}{S(t)}$$

e poiché ***f(t)*** non è altro che la derivata rispetto a ***t*** di ***F(t)***, l'equazione diventa[2]:

$$h(t) = \frac{dF(t)}{dt} \frac{1}{S(t)} = \frac{f(t)}{S(t)} \qquad \text{(4a)}$$

da cui segue

$$h(t) = -\frac{d}{dt} \log S(t)$$

e dunque

$$S(t) = \exp\left(-\int_0^t h(u)du \right) = \exp(-H(t)) \qquad \text{(5)}$$

Dalla stima del rischio ***h(t)*** e della corrispondente funzione di sopravvivenza si ottiene una stima dei tempi di accadimento dell'evento che è oggetto dell'analisi.

Una misura di sintesi comunemente utilizzata per rappresentare la sopravvivenza di un gruppo di individui è la durata mediana[3].

[2] La funzione di densità di probabilità, ***f(t)*** è definita come $f(t) = \lim_{\delta t \to 0} \Pr\{t \le T < t + \delta t\}$; dalla (4a) si osserva che, note due delle tre funzioni fondamentali, è possibile calcolare la terza.

[3] Nel caso di modelli di durata, in cui si considerano anche osservazioni censurate oltre a durate effettive non censurate, è più appropriato utilizzare la durata mediana invece della durata media. Infatti, la durata media di sopravvivenza stimata non sempre è calcolabile nei modelli non parametrici, nel caso in cui sia particolarmente elevata la frequenza di durate censurate

4 STIMA NON PARAMETRICA DELLA CURVA DI SOPRAVVIVENZA: IL METODO DI KAPLAN-MEIER

E' possibile analizzare preliminarmente la curva di sopravvivenza ricorrendo a tecniche non parametriche, che non richiedono la formulazione di ipotesi sulla distribuzione di probabilità della variabile T.

Uno dei metodi utilizzati è stato proposto da **Kaplan–Meier**. Se indichiamo con $t_1, t_2, t_3 \ldots t_k, \ldots$ in ordine successivo i tempi in cui si verifica un evento, gli intervalli compresi tra durate successive t_j - t_{j+1} includeranno almeno un evento.

Indicando con n_j gli individui che ancora non hanno sperimentato eventi al tempo t_j e con e_j gli eventi che si verificheranno tra t_j e t_{j+1}, la probabilità stimata di sopravvivenza nell'intervallo *j*-simo è data da:

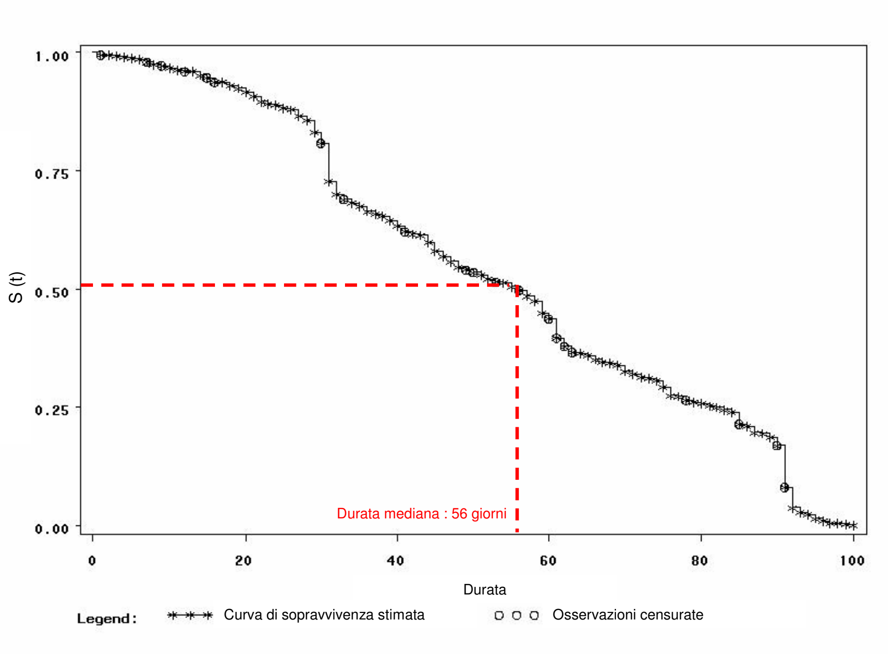
$$\hat{S}(t_j) = \frac{n_j - e_j}{n_j} \qquad (6)$$

Assumendo che gli eventi ***e*** degli individui osservati siano tra loro indipendenti, la funzione di sopravvivenza ***S(t)*** può essere stimata come:

$$\hat{S}(t) = \prod_{j=1}^{k} \frac{n_j - e_j}{n_j} \qquad (7)$$

A partire dalle eq (5) e (6) è possibile ottenere una stima non parametrica di ***H(t)*** (cfr. Collett 1994).

Figura 4 - Esempio di curva di sopravvivenza stimata (Kaplan–Meier) per la fase di progettazione preliminare (selezione per durate <=100 gg)



La stima non parametrica della curva di sopravvivenza è utile per analizzare un gruppo omogeneo di individui, per esempio per verificare l'esistenza di particolari relazioni tra rischio ***h(t)*** e tempo ***t***, oppure per confrontare due o più gruppi di individui, classificati rispetto a diverse modalità di una stessa variabile. Come verrà illustrato più avanti[4], nella nostra applicazione l'analisi non parametrica preliminare è stata svolta per la scelta della distribuzione di probabilità che meglio approssimava la distribuzione osservata della variabile T.

Per analizzare la sopravvivenza in funzione di un **insieme** di caratteristiche, si ricorre a modelli statistici più complessi.

5 ANALISI PARAMETRICA

Utilizzando modelli parametrici di sopravvivenza è possibile studiare in che modo i tempi di attesa fino al verificarsi di un evento di interesse siano influenzati da una o più caratteristiche degli individui.

[4] Cfr. Par. 5.1

Diversamente dai modelli di regressione lineari 'classici', in cui ciò che viene modellizzato in funzione di un set di variabili esplicative è la ***media*** della variabile dipendente di interesse, nei modelli di sopravvivenza la variabile dipendente è rappresentata dal rischio ***h(t)***:

$$h(t) = g(x;\beta;t) \tag{8}$$

dove x indica un vettore di variabili esplicative, β i parametri corrispondenti e ***t*** il tempo.

A partire dalla stima del rischio, $\hat{h}(t)$ ottenuta per ogni singolo individuo, è possibile poi ottenere una stima della curva di sopravvivenza, $\hat{S}(t)$ e di conseguenza, la stima della durata mediana.

Assumendo

$$g(x;\beta;t) = \exp(x;\beta;t)$$

si assume che il **logaritmo** della funzione di rischio ***h(t)*** sia una funzione del tempo ***t*** e di una serie di variabili esplicative[5] $\mathbf{X_1 \ldots X_{n-1}}$, che rappresentano le caratteristiche degli interventi[6]:

$$\log[h(t)] = l(t, x_1, \ldots, x_n) = \beta_0 + \beta_1 x_1 + \ldots + \beta_{n-1} x_{n-1} + \beta_n f(t) \tag{9}$$

da cui possiamo riscrivere la (8) come:

---

[5] Il parametro generico β_j o più precisamente ***exp(β))*** misura l'effetto della alla variabile X_j sul rischio *h(t)* in termini relativi rispetto al rischio base.

A titolo esemplificativo, se consideriamo il modello semplificato, con una sola variabile esplicativa X; dicotomica, ad esempio:

X= Macro Area (0= Centro – Nord; 1= Mezzogiorno)

allora si avrà che:

$$\frac{h(t \mid MacroArea = 1)}{h(t \mid MacroArea = 0)} = \frac{h_0(t)e^{\beta}}{h_0(t)e^{0}} = e^{\beta}$$

[6] I modelli di durata consentono di inserire oltre a variabili 'statiche', anche variabili tempo – dipendenti.

$$h_i(t) = h_0(t)e^{\beta_1 x_{1i} + ... + \beta_{n-1} x_{n-1i}}$$

dove $h_0(t)$ indica la funzione di rischio base (baseline hazard function)[7].

A particolari ipotesi sulla dipendenza del rischio ***h(t)*** dal tempo corrispondono assunzioni sulla distribuzione di probabilità della variabile casuale tempo T[8].

Si parla in questo caso di modelli **parametrici**, e le distribuzioni più comuni utilizzate in questo tipo di analisi sono :

- la **esponenziale**: hp. rischio costante in funzione del tempo
$$\log[h(t)] = \beta_0 + \beta_1 x_1 + ... + \beta_{n-1} x_{n-1}$$
- la **Weibull**: hp. rischio funzione monotona rispetto al tempo
$$\log[h(t)] = \beta_0 + \beta_1 x_1 + ... + \beta_{n-1} x_{n-1} + \beta_n \log t$$

La formulazione adottata fino a questo momento è quella più comunemente utilizzata in letteratura e nella maggior parte dei software che permettono di stimare modelli di durata. Tuttavia, i modelli qui presentati possono essere formulati in modo del tutto equivalente rispetto alla durata T, invece che rispetto al rischio, $h(t)$[9].

Si può, quindi, scrivere:

$$\log T = \alpha_0 + \alpha_1 x_1 + ... + \alpha_n x_n + \sigma\varepsilon \tag{10}$$

Questa formulazione, che è stata utilizzata nei nostri modelli, stimati con il pacchetto statistico SAS, non permette di stimare direttamente un valore di T, ma i parametri della sua distribuzione.

Gli α_i sono i parametri incogniti associati alle variabili esplicative X_i[10], α_0 è l'intercetta e σ è il parametro associato alla variabile casuale ε, che coglie l'allontanamento del valore di *log T* dalla parte lineare del modello.

---

[7] Il baseline hazard $h_0(t)$ si ottiene ponendo a 0 le variabili esplicative X, esso dunque rappresenta il rischio calcolato in assenza di covariate.
Nel caso in cui le variabili esplicative siano variabili dummy (dummy coding), e dunque il valore '0' è associato ad una modalità scelta come riferimento, il baseline hazard si riferisce al generico individuo caratterizzato dalle modalità di riferimento delle covariate considerate.

[8] Cox (1972) ha proposto un modello 'semiparametrico' alternativo che non presuppone alcuna assunzione per la distribuzione di probabilità della variabile tempo T.

[9] Cfr. Collett (1994), Allison (1995).

[10] Esiste una relazione diretta tra i parametri β_i della (8) e gli α_i della (9), che dipende dalla distribuzione assunta per la T. In generale, se β_i misura l'effetto della variabile Xi sulla funzione di rischio e α_i l'effetto della stessa variabile sul tempo trascorso fino al verificarsi dell'evento, a valori elevati del primo,

Per ogni specifica combinazione di variabili esplicative $X_1, X_{2,}......, X_{n.-1}$ osservata su un progetto, a partire dai parametri stimati nella (9) è possibile ottenere una distribuzione di probabilità della durata T, a partire dalla quale è possibile calcolare la <u>durata mediana</u> della fase associata a quello specifico progetto e tutti i percentili della distribuzione

5.1 Stima dei parametri: il metodo di massima verosimiglianza

La stima dei parametri può essere ottenuta con il metodo della **massima verosimiglianza**. E' dunque necessario definire la funzione di verosimiglianza del campione di individui osservato, che rappresenta la probabilità congiunta dei casi osservati, calcolata in funzione dei parametri incogniti del modello.

Nella costruzione della funzione di verosimiglianza, è necessario distinguere tra casi censurati e non censurati.

Partiamo tuttavia dall'ipotesi che non ci siano casi censurati, ma solo durate concluse, t_i .

Per un generico campione di n individui, che si assumono indipendenti tra di loro, la funzione di verosimiglianza (Likelihood) è definita dal prodotto delle probabilità dei singoli:

$$L = \prod_{i=1}^{n} f_i(t_i) \qquad (11)$$

dove f_i è la funzione di densità di probabilità dell'individuo i-esimo, che dipende dalle covariate specifiche dell'individuo (cfr. (2)).

Nel caso in cui nel campione abbiamo durate censurate, la (11) si modifica; infatti, nel caso in cui la durata sia censurata, tutto quello che possiamo dire è che l'evento, per questi individui, si verificherà in un tempo t'_i successivo al massimo osservato. Ma la probabilità che il tempo di accadimento di un evento sia maggiore di un certo valore osservato, t_i è dato dalla funzione di sopravvivenza $S(t_i)$

Pertanto, se per semplicità ipotizziamo che i primi *r* individui hanno sperimentato l'evento, e dunque registrano durate concluse, e i successivi $n - r$ individui hanno durate censurate, possiamo riscrivere la (11) come

$$L = \prod_{i=1}^{r} f_i(t_i) \prod_{r+1=1}^{n} S_i(t_i) \qquad (12)$$

corrispondenti ad una elevata probabilità che l'evento accada, corrisponderanno valori contenuti del secondo, associati a tempi rapidi di accadimento. Cfr. Collett (1994), Allison (1995).

Se indichiamo con δ_i la variabile dummy, che vale 1 se la durata è conclusa, 0 se è censurata, la (12) diventa:

$$L = \prod_{i=1}^{n} f_i(t_i)^{\delta_i} S_i(t_i)^{1-\delta_i} \qquad (12a)$$

Una volta scelta la distribuzione per la variabile T, è possibile sostituire la $f_i(t_i)$ e la $S_i(t_i)$ con le relative espressioni.

Nel caso più semplice, in cui si assume per la variabile T la distribuzione esponenziale, essendo

$$f_i(t_i) = \lambda_i \exp(-\lambda_i t_i)$$

e

$$S_i(t_i) = \exp(-\lambda_i t_i)$$

la (12a) diventa

$$L = \prod_{i=1}^{n} [\lambda_i \exp(-\lambda_i t_i)]^{\delta_i} \exp(-\lambda_i t_i)^{1-\delta_i} = \lambda_i^{\delta_i} \exp(-\lambda_i t_i)$$

dove $\lambda_i = \exp(-\beta x)$.

6 SCELTA DEI MODELLI PER LE CINQUE FASI

Per ogni fase è stato stimato un modello specifico: sono state condotte analisi preliminari, sia per la valutazione della qualità dei dati disponibili sulle durate per ogni fase, complete o censurate, sia per la osservazione delle caratteristiche dei progetti con durate 'valide' e la scelta della forma funzionale appropriata

L'ipotesi alla base dei 5 modelli stimati è che la durata per il completamento delle fasi procedurali di un progetto di investimento sia influenzata sia dalle caratteristiche strutturali dei progetti (**settore, tipologia di intervento, costo complessivo, tipologia di ente attuatore,** della **fonte di finanziamento**) sia da un insieme di fattori legati al territorio in cui tali interventi vengono realizzati[11].

[11] Cfr. I tempi di attuazione delle Opere Pubbliche, Tab. 1

Nella stima di un modello di durata parametrico, un primo passo fondamentale è l'osservazione dei dati di durata disponibili, allo scopo di individuare la distribuzione più appropriata a descrivere la relazione tra rischio *h(t)* e tempo t.

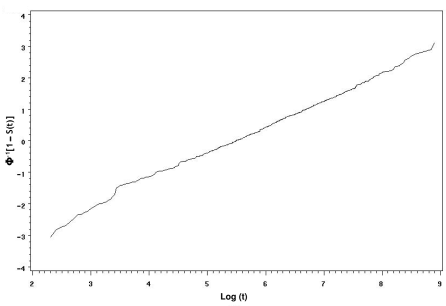
Per questo, è stata condotta preliminarmente un'analisi non parametrica dei dati ed una analisi grafica (cfr. Fig. 3).

Per quanto riguarda la progettazione preliminare, definitiva ed esecutiva, così come per la durata del bando, per la T si è assunta una distribuzione Log Normale, che corrisponde ad una distribuzione del rischio *h(t)* non-monotona, con andamento ad 'U' rovesciata[12].

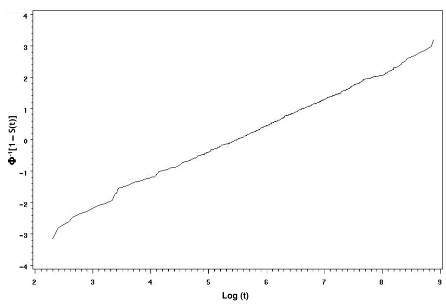
Tale scelta è stata effettuata sulla base dell'analisi della relazione tra la funzione di sopravvivenza osservata, $S(t)$, ed il tempo t. Se l'ipotesi di una distribuzione Log-Normale per la T è valida, il grafico della funzione di ripartizione della Normale inversa, $\Phi^{-1}[1-\hat{S}(t)]$ rispetto a $\log(t)$ approssimerà una retta; come si vede dal grafico qui sotto riportato, ciò si verifica nel caso delle durate della progettazione preliminare, definitiva ed esecutiva, e per l'aggiudicazione del bando.

Figura 5 - Valutazione grafica della forma funzionale della variabile casuale T = durata della fase. Progettazione preliminare, definitiva, esecutiva e bando.

Progettazione preliminare

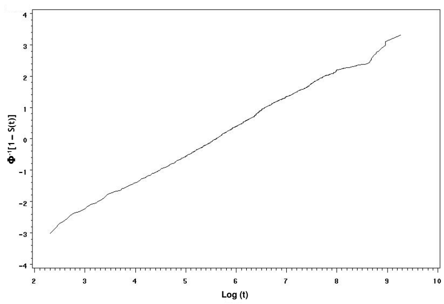


Progettazione definitiva

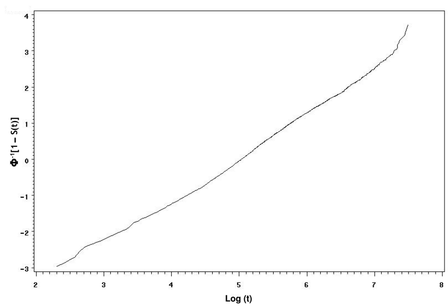


[12] Il rischio è 0 per t=0, cresce fino ad un punto di massimo e poi decresce nuovamente, per tendere a 0 per t che tende ad ∞.

Progettazione esecutiva



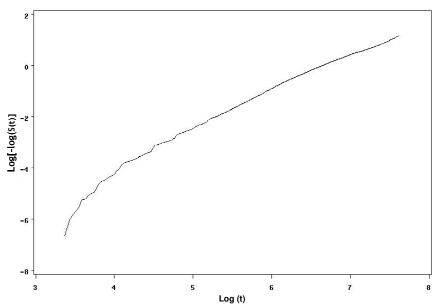
Aggiudicazione bando



Nel caso in cui la variabile T si distribuisca, invece, secondo una Weibull, allora la relazione tra Log(-log(S(t)) e il logaritmo della durata, t, sarà lineare. Questa ipotesi risulta accettabile per la durata dell'esecuzione dei lavori, come dimostra il grafico sottostante.

Figura 6 - Valutazione grafica della forma funzionale della variabile casuale T = durata della fase. Esecuzione lavori.

Esecuzione lavori



Di seguito vengono riportati i parametri stimati per i modelli relativi alle 5 fasi considerate, e, fase per fase, illustrati i risultati più rilevanti, con particolare riferimento alle variabili categoriche considerate.

Come già accennato precedentemente[13], l'esponenziale del parametro **β** indica l'effetto relativo della modalità a cui è associato, **rispetto alla modalità di riferimento**, per cui tale valore è pari ad 1, al netto di tutte le altre variabili considerate nel modello.

La scelta della modalità è arbitraria.

[13] Cfr. nota 5.

Tabella 1 – Parametri dei modelli di durata per le cinque fasi procedurali - continua

VARIABILI	MODALITA'	FASI									
		Prog. Preliminare		Prog. Definitiva		Prog. Esecutiva		Aggiudicazione Bando		Esecuzione Lavori	
		exp(β)		exp(β)		exp(β)		exp(β)		exp(β)	
INTERCETTA		2.02	*	0.24	***	15.59	***	11.10	***	0.04	***
REGIONE	ABRUZZO	0.73	***	0.57	***	0.69	***	0.93	**	1.07	
	BASILICATA	1.00		0.34	***	1.56	***	1.22	***	1.35	***
	CALABRIA	0.74	*	0.80		0.85		0.98		1.09	
	CAMPANIA	0.74	**	0.80	**	1.15		1.25	***	1.04	
	EMILIA-ROMAGNA	1.01		0.58	***	0.53	***	0.86	***	0.87	***
	FRIULI-VENEZIA GIULIA	0.64	***	1.10		0.40	***	1.07	***	0.76	***
	LAZIO	0.58	***	0.66	***	1.10		1.20	***	1.11	**
	LIGURIA	1.27	*	0.75	***	0.65	***	1.06	*	1.09	
	LOMBARDIA	0.88		0.68	***	0.70	***	1.03		0.96	
	MARCHE	1.50	**	0.60	***	0.87		1.00		0.77	***
	MOLISE	0.43	***	0.65	***	1.11		0.99		0.88	**
	PIEMONTE	1.08		0.78	***	0.47	***	1.04		0.76	***
	PUGLIA	0.60	***	0.64	***	1.12		1.16	***	1.04	
	SARDEGNA	0.83		0.71	***	1.40	***	1.07	**	0.77	***
	SICILIA	0.73	**	1.65	***	2.19	***	1.20	***	0.80	***
	TOSCANA	1.06		0.80	**	1.00		1.19	***	1.18	***
	TRENTINO-ALTO ADIGE	1.10		0.71	***	0.75	***	1.01		0.91	*
	UMBRIA	1.00		1.00		1.00		1.00		1.00	
	VALLE D'AOSTA	1.08		0.78	***	0.47	***	1.67	***	0.76	***
	VENETO	1.10		0.71	***	0.75	***	1.07	***	0.91	*
COSTO(Log-log)		6.06	***	16.05	***	3.17	***	2.40	***	41.08	***
SETTORE	ALTRI TRASPORTI	0.91		0.66	***	0.84	***	1.15	***	1.00	
	AMBIENTE	1.13	*	0.87	**	0.90	**	0.94		1.12	***
	CICLO INTEGR. DELL'ACQUA	0.96		0.87	**	1.15	***	0.96	**	1.20	***
	VARIE	1.12	*	0.92		1.08	**	1.04	***	1.23	***
	VIABILITÀ	1.00		1.00		1.00		1.00		1.00	
ENTE ATTUATORE	COMUNI	1.13	***	0.96		0.79	***	0.92	***	1.00	
	PROVINCE	1.01		1.07		0.98		1.11	***	0.89	
	REGIONI	0.96		0.99		1.04		1.24	***	0.98	***
	GESTIONE RETI	1.21	***	1.17	***	1.25	***	0.96	**	0.92	
	MINISTERI	1.36	***	0.84	**	1.08		1.61	***	1.23	*

	ALTRI ENTI	1.00	1.00	1.00	1.00	1.00 ***

Tabella 1 – Parametri dei modelli di durata per le cinque fasi procedurali - segue

VARIABILI	MODALITA'	FASI				
		Prog. Preliminare	Prog. Definitiva	Prog. Esecutiva	Aggiudicazione Bando	Esecuzione Lavori
		exp(β)	exp(β)	exp(β)	exp(β)	exp(β)
TIPOLOGIA DI INTERVENTO	NUOVA REALIZZAZIONE	1.13 ***	1.26 ***	1.07 ***	1.10 ***	1.00
	REST./RECUP./MANUT.	1.00	1.00	1.00	1.00	1.00
PROCEDURA DI SELEZIONE	APERTA				1.29 ***	
	RISTRETTA				1.85 ***	
	NEGOZIATA				1.00	
INDICATORE DI SVILUPPO ECONOMICO		0.93 **	0.95 *	1.09 ***	1.00	
INDICATORE DI QUALITA' DELL'AMBIENTE		0.93 **	1.00	1.05 **	0.97 **	1.00
INDICATORE DI SVILUPPO SOCIALE			1.00	1.00	1.00	***
INDICATORE DI CRIMINALITA' ORGANIZZATA			0.86 ***	0.94 **	1.05 ***	
INDICATORE DI CRIM. ' CONTRO LA PERSONA						1.04 *
DISTRIBUZIONE DI PROBABILITA'		Log-normale	Log- normale	Log- normale	Log- normale	Weibull

Dalla lettura dei parametri, è possibile sintetizzare alcuni risultati principali rispetto alle variabili considerate.

Per quanto riguarda l'effetto del **territorio**, che nei modelli è colto dalla variabile categorica regionale, è interessante sottolineare due aspetti:

1. nelle singole regioni considerate, non si riscontrano comportamenti 'coerenti' tra le diverse fasi; si osserva invece che a durate mediamente brevi in alcune fasi corrispondono tempi mediamente più lunghi in altre, e viceversa;
2. non emerge una netta dicotomia Nord / Sud : tempi brevi o lunghi nelle varie fasi si rilevano tanto nelle regioni del Centro Nord, quanto in quelle del Mezzogiorno.

Analogamente, per quanto riguarda il **settore** e la **tipologia di ente attuatore** non è possibile individuare alcuna categoria con durate sistematicamente brevi o lunghe in tutte le fasi considerate. Ad esempio, un intervento nel settore del "Ciclo integrato dell'acqua" registrerà tempi più brevi nel completare la progettazione preliminare e la definitiva, così come sarà più veloce in fase di aggiudicazione, rispetto alla categoria di riferimento "Altri Trasporti", ma impiegherà tempi più lunghi del 15% in fase di progettazione esecutiva, e del 20% più lunghi in fase di esecuzione lavori.

Per quanto riguarda la **tipologia di intervento** si registrano tempi generalmente più lunghi per interventi di nuova realizzazione, rispetto a lavori di ristrutturazione, restauro o manutenzione.

L'interpretazione dei parametri relativi agli indicatori di contesto **socio economico** non è immediata: se, infatti, da una parte migliori condizioni economiche ed ambientali sono associate a durate più brevi in fase di progettazione preliminare e definitiva, e in fase di aggiudicazione, come indica il parametro inferiore a 1, dall'altra corrispondono a tempi mediamente lunghi di progettazione esecutiva. Una possibile spiegazione di questo risultato è che una maggiore efficienza delle amministrazioni delle zone economicamente più sviluppate si traduce in maggiore celerità nel completare le prime due progettazioni, che principalmente prevedono il disbrigo di pratiche burocratiche e la richiesta di permessi e autorizzazioni, mentre corrisponde a tempi più lunghi nella definizione dei dettagli fondamentali del progetto, in termini di costi e tipo di lavori previsti. Quindi, maggiore lentezza nel completamento dei alcune fasi non necessariamente può essere interpretato come segnale negativo, ma può essere letto come indice di maggiore accuratezza ed efficienza. Analogamente, per quanto riguarda gli indici di Criminalità i risultati possono apparire inizialmente controintuitivi: valori elevati degli indicatori di diffusione della Criminalità organizzata e di Criminalità contro le persone sono associate con migliori performance in fase di progettazione; tuttavia, essi si traducono poi in tempi più lenti in fase di aggiudicazione e di esecuzione lavori. Da una lettura congiunta dei parametri dei cinque modelli emerge quindi che lì dove la criminalità è più diffusa, la progettazione viene completata più velocemente (e, forse, in maniera più sbrigativa e meno precisa), mentre l'aggiudicazione e l'esecuzione dei lavori richiede più tempo, a causa di interruzioni e varianti rispetto al progetto originario, che spesso si traducono in un aumento dei costi.

BIBLIOGRAFIA

Allison P. (1995), Survival Analysis Using the SAS® System: A Practical Guide, SAS Institute Inc., Cary NC.

Collett D. (1994), Modelling survival data in medical research, Chapman & Hall, London

Cox D.R. (1972), Regression Models and Life Tables (with discussion), Journal of the Royal Statistical Society (B) 34: 187 – 220