

I MODELLI LOG-LINEARI NELLA RICERCA SOCIOLOGICA

di GIORGIO CHIARI - PIERANGELO PERI



QUADERNO 11/12
Dipartimento di Politica Sociale
Università di Trento

I Quaderni del Dipartimento di Politica Sociale costituiscono una iniziativa editoriale finalizzata alla tempestiva diffusione in ambito universitario di *materiale di ricerca, riflessioni teoriche e resoconti* di Seminari di studio di particolare rilevanza. L'accettazione dei diversi contributi è subordinata all'approvazione di un'apposita Commissione scientifica, che si avvale anche del parere di *referees* esterni al Dipartimento. Responsabile della serie dei Quaderni: Prof. Marino Livolsi, Ordinario di Sociologia, Trento. Le richieste dei Quaderni vanno inviate ai rispettivi autori.

1. E. BAUMGARTNER, *L'identità nel cambiamento*, 1983.
2. C. SARACENO, *Changing the Gender Structure of Family Organization*, 1984.
3. G. SARCHIELLI, M. DEPOLO e G. AVEZZU', *Rappresentazioni del lavoro e identità sociale in un gruppo di lavoratori irregolari*, 1984.
4. S. GHERARDI, A. STRATI, (a cura di), *Sviluppo e declino: La dimensione temporale nello studio delle organizzazioni*, 1984.
- 5/6 A. STRATI (ed.), *The Symbolics of Skill*, 1985.
7. G. CHIARI, *Guida bibliografica alle Tecniche di ricerca sociale*, 1986.
8. M. DEPOLO, R. FASOL, F. FRACCAROLI, G. SARCHIELLI, *L'azione negoziale*, 1986.
9. C. SARACENO, *Corso della vita e approccio biografico*, 1986
10. R. PORRO (a cura di), *Le comunicazioni di massa*, 1987.

Dipartimento di Politica Sociale
 Università di Trento
 via Verdi 26, I-38100 Trento
 Tel. 0461-881111/881322
 telex 400674 UNIVTN I
 Italia

Composizione: Copisteria Paradisi Trento
 © Copyright Dipartimento di Politica Sociale, 1987

Indice

Premessa	pag. 5
Introduzione: le origini dei modelli log-lineari	pag. 7
Cap. 1 - Le varie forme di modelli log-lineari	pag. 15
1.1. Il chi quadrato come misura della bontà di adattamento fra valori osservati e valori teorici	pag. 17
1.2. Modelli log-lineari per tabelle 2x2	pag. 28
1.2.1. Forma moltiplicativa del modello	pag. 38
1.2.2. Forma additiva del modello	pag. 50
1.2.2.1. Modelli additivi sulle log-probabilità	pag. 54
1.2.2.2. Modelli additivi sulle log-frequenze	pag. 55
1.3. Corrispondenza con modelli additivi e moltiplicativi	pag. 63
1.4. Altri modelli per tabelle 2x2: modelli saturi	pag. 67
Cap. 2 - Analisi di tabelle multidimensionali	pag. 73
2.1. La tabella generale a tre dimensioni con variabili dicotomiche	pag. 75
2.2. Il problema dei segni	pag. 85
2.3. Il calcolo dei parametri nel caso di tabelle 2x2x2	pag. 88
2.4. Modelli per tabelle a quattro dimensioni con variabili dicotomiche (2x2x2x2)	pag. 98
2.5. Modelli log-lineari con variabili politomiche	pag. 106
Cap. 3 - I modelli logit	pag. 117
3.1. Un modello logit per l'interpretazione del proseguimento degli studi in Provincia di Trento	pag. 124
3.1.1. Il modello generale	pag. 124
3.1.2. I modelli per i diversi indirizzi di studi	pag. 128
Cap. 4 - Modelli log-lineari ed analisi causale	pag. 135
Conclusioni	pag. 147
Bibliografia	pag. 151
Appendice	pag. 175
Allegati	pag. 215

Premessa

Quando, nel 1984, stimolati da A. de Lillo, decidemmo di studiare in modo approfondito le potenzialità offerte dai modelli log-lineari nell'analisi delle variabili qualitative, ci scontrammo immediatamente con una serie di difficoltà. La letteratura di base ci lasciava in molti casi con seri dubbi e molte lacune data la grande quantità di "passaggi" saltati nei calcoli e la tendenza a dare per scontate cose che, per chi non è uno statistico puro, scontate non sono. Allo stesso modo gli esempi riportati, in larga parte costruiti ad arte e perciò lontani dalla realtà quotidiana del ricercatore, sempre diversi e a volte fuorvianti, rendevano ardua l'impresa di comprendere fino in fondo il percorso di analisi e gli assunti sottostanti a questi modelli. Non ci sembrava sufficiente infatti essere in grado di interpretare i dati forniti dal computer utilizzando il programma ECTA di Goodman, ma volevamo "dominare" tutto il percorso di calcolo che portava ai risultati finali. Fu così che iniziammo un lungo e faticoso lavoro di studio per cercare di ricostruire l'intero mosaico dell'analisi rintracciandone i vari tasselli attraverso la disamina dei numerosi e non sempre omogenei contributi esistenti sull'argomento.

Arrivammo così alla decisione di evitare ad altri molte delle difficoltà da noi incontrate, attraverso questo saggio che ripercorre, in un modo che speriamo comprensibile, l'iter del nostro processo di comprensione dei modelli log-lineari, logit e di analisi causale con variabili qualitative. Ci ha guidato, nella stesura, una costante preoccupazione didattica che ci ha portato forse ad eccedere nella proposizione di calcoli (memori del tempo e della fatica impiegati per rifare manualmente quelli relativi agli esempi riportati da altri) con il rischio di essere in alcuni casi ridondanti, come gli addetti ai lavori non mancheranno di farci rilevare. I destinatari di questo lavoro dovrebbero essere ricercatori e studenti di sociologia, ma anche

psicologi sociali, statistici sociali (junior) ecc. con una discreta conoscenza della statistica di base.

Vogliamo anche ringraziare A. de Lillo e H. Schadee che, con i loro consigli e critiche ci hanno aiutato in questa impresa; B. Bazzanella, direttore della biblioteca dell'Università, al quale siamo debitori per l'aiuto prezioso fornitoci nel rintracciare la corposa bibliografia e, infine, tutti coloro che abbiamo trascurato nel corso delle nostre vacanze-studio.

Per concludere, dovremmo 'dividerci' la stesura finale; operazione difficile dato che quasi ogni parola del testo, ogni calcolo, ogni esempio, è stato pensato, scritto e riscritto con il contributo di entrambi gli autori che si assumono solidalmente la responsabilità dell'intero testo. Non potendo però esimerci da una divisione formale, l'introduzione, i capitoli 2, 3 e l'appendice sono dovuti a G. Chiari; i capitoli 1 e 4 sono dovuti a P. Peri.

*Giorgio Chiari
Pierangelo Peri*

Lago di Iseo, luglio 1987

Introduzione

LE ORIGINI DEI MODELLI LOG-LINEARI

I modelli log-lineari, assieme ad altre tecniche parametriche per l'analisi dei dati qualitativi, appartengono ad una famiglia che discende, pur con alcune significative variazioni, dal modello lineare generalizzato. Si tratta di una vera e propria paternità che trova una corrispondenza, per il trattamento dei dati quantitativi, nella base fornita dal modello generale lineare per lo sviluppo dei modelli di regressione e per l'analisi della varianza.

Quando una delle variabili di una tavola di contingenza è trattata come variabile dipendente queste tecniche parametriche (di tipo lineare generalizzato) risultano estremamente simili alla regressione ordinaria sia per quanto riguarda la logica dell'analisi, sia per quanto riguarda le procedure di stima.

Per variabili di tipo qualitativo, considerato il livello di misurazione dei dati che renderebbe inapplicabile un modello di regressione, è necessario individuare una strategia che renda compatibile l'uso di questa tecnica con la configurazione dei dati. Questa strategia sta proprio nell'operare una sorta di analisi di regressione ponderata su alcune trasformazioni delle frequenze di cella in una tavola di contingenza, predicendo le frequenze così trasformate come funzione di punteggi assegnati a categorie delle variabili indipendenti (cfr. Goodman, 1972; Haberman, 1968; Johnston, 1972).

Come vedremo, i tre modelli alternativi (linear probability (LPM), logit, log-linear) si assomigliano notevolmente e utilizzano, con alcune variazioni, le stesse procedure della regressione ordinaria.

Come è noto la regressione ordinaria è il modello di analisi più idoneo quando si voglia considerare una variabile dipendente funzione di numerose variabili indipendenti. Qualora le variabili indipendenti siano di tipo categoriale (o qualitativo) si possono applicare le normali tecniche di regressione ordinaria (OLS) attraverso una semplice trasformazione di esse

con la nota procedura di 'dummy coding' che ne mantiene inviolati gli assunti fondamentali.

Si parla, a questo proposito, di equivalenza tra analisi di regressione, analisi della varianza (Fennessey, 1968) e analisi di classificazione multipla (Andrews et al., 1967). Viceversa, quando la variabile dipendente è categoriale, non è possibile applicare le procedure di stima della regressione ordinaria (OLS)¹

Questa difficoltà viene in parte superata nel caso di variabili dipendenti dicotomiche. In questo caso, assai frequente nella ricerca sociologica, i valori della variabile dipendente vengono trattati in modo continuo nell'intervallo di probabilità (o di proporzione) 0-1². Poiché la media di una variabile dicotomica è interpretabile come proporzione (o probabilità) è ragionevole applicare alla tavola di contingenza un modello di analisi di scomposizione lineare simile alla regressione (*regression-like*, Swafford, 1980, p. 665).

In questo caso l'interpretazione dei coefficienti stimati riguarderà come le variabili indipendenti influenzano la probabilità che i soggetti cadano in una data categoria-cella.

Rimane tuttavia il problema, cruciale dal punto di vista statistico, che una variabile dicotomica tipicamente viola due degli assunti fondamentali della stima dei minimi quadrati, vale a dire omoschedasticità (varianza costante) e distribuzione normale della variabile dipendente ai vari livelli delle variabili indipendenti. Ignorare tali assunti significa inficiare sia l'efficienza delle stime che la validità dei test di significatività.

Tale problema è stato risolto, già alla fine degli anni '60, da Grizzle, Starmer e Koch che hanno sviluppato alcune modifiche alle procedure di

¹ La regressione ordinaria genera infatti previsioni numeriche su una scala a intervalli o di rapporti che non hanno alcun significato se riferite a una variabile dipendente categoriale.

² Ad es. se la variabile dipendente analizzata è la continuazione degli studi e la proporzione di coloro che continuano dopo la 3^a media è di 4197 su 6451 ($p = .651$), la variabile "continuazione degli studi" assumerà valore 1 per i 4197 soggetti che continuano e 0 per i 2254 ($p = .349$) che non continuano. La media ponderata dei valori così assegnati corrisponde appunto alla proporzione $p = .651$.

stima OLS che permettono di procedere in modo legittimo anche quando tali assunti vengono a mancare. La procedura GSK (dai nomi degli autori) da essi elaborata fa uso di una routine di minimi quadrati generalizzati ponderati che sostituisce al tradizionale test di significatività F un chi-quadrato statisticamente più efficiente. In questo modo l'analisi di regressione procede in maniera analoga a quella su variabili indipendenti continue così come l'interpretazione dei coefficienti stimati (Swafford, 1980, p. 666).

Delineata in tal modo la struttura di base dei modelli parametrici per l'analisi dei dati qualitativi vediamo più da vicino le principali somiglianze e differenze esistenti fra i tre tipi di modelli che consentono l'analisi quantitativa delle frequenze multivariate, vale a dire quelli di probabilità lineare, logit e log-lineari.

Un primo elemento di differenziazione sta nella "variabile criterio", definita come variabile dipendente dei modelli. I modelli di probabilità lineare assumono come variabile criterio la proporzione di soggetti che si trovano in una data categoria-cella di una variabile dipendente dicotomica. I modelli logit assumono come variabile criterio il logaritmo degli odds vale a dire dei rapporti fra le frequenze di soggetti che si trovano in due categorie dicotomiche mutualmente esclusive.

Nei modelli log-lineari la variabile criterio è rappresentata direttamente dalle frequenze (versione moltiplicativa) o dalle log-frequenze (versione additiva).³

Un secondo elemento di differenziazione fra i vari modelli riguarda i "metodi di stima" dei coefficienti. Nelle prime due classi di modelli (di probabilità lineare e logit) la stima dei coefficienti avviene secondo il metodo dei minimi quadrati (GLS) testati con la tecnica GSK prima ricordata; i modelli log-lineari applicano il metodo della massima verosimiglianza (ML). Questa differenza, apprezzabile qualora le dimensioni campionarie siano ridotte, tende a ridimensionarsi quando i

³ Come vedremo nel corso del lavoro esistono altre due versioni di modelli log lineari in cui le variabili criterio risultano, rispettivamente, le probabilità e le log-probabilità di una frequenza di cella attesa.

campioni sono consistenti. In tal caso le stime GLS e ML tendono a produrre risultati sostanzialmente equivalenti e ad assumere le stesse proprietà 'Best Asymptotic Normal' (BAN) (Swafford, 1980, p. 679).

Un ulteriore elemento di differenziazione contrappone i modelli che definiscono esplicitamente una variabile dipendente (LPM e logit) e i modelli log-lineari generali nei quali non si assume alcuna relazione causale fra le variabili che risultano così tutte predittive della configurazione delle frequenze di cella.

Per quanto riguarda il confronto fra modelli di probabilità e modelli logit, i più adatti alla spiegazione di variabili sociologiche 'dipendenti', esistono elementi a favore di entrambi. I modelli probabilistici, se da un lato presentano una più diretta ed immediata interpretabilità dei coefficienti stimati, dall'altro possono produrre stime 'impossibili' (negative o superiori all'unità) il che non avviene nei modelli log-lineari. Questi ultimi sembrano preferibili quando si lavora con variabili in cui le proporzioni osservate si avvicinano ai valori estremi (0 e 1). In questo caso i modelli logit sarebbero più adatti a trattare tali effetti marginali (Theil, 1970).

In ogni caso modelli probabilistici lineari e modelli logit, pur essendo sostanzialmente equivalenti, non sono identici. Essi conducono agli stessi risultati soltanto se le variabili criterio sono fra loro altamente correlate (Swafford, 1980, p. 678).

Mentre i modelli di regressione vengono solitamente utilizzati per analizzare e predire punteggi di singoli individui di un campione, i modelli log-lineari operano su configurazioni delle frequenze di cella o delle relative proporzioni (trasformate in logaritmi) nella tavola di contingenza.

Per concludere queste brevi note introduttive, vale la pena di rivisitare sinteticamente i diversi modelli di analisi log-lineare presenti in letteratura, tutti riconducibili alle tre classi di modelli precedentemente illustrati, vale a dire modelli di probabilità lineare, modelli logit e modelli log-lineari.

Alla prima classe dei modelli di probabilità lineare appartengono sia il 'D-Systems' (Davis, 1976; Ammassari, 1986), sia il 'Weighted Regression Model for Proportions' (Grizzle et al., 1969; Taylor, 1983) che sono sostanzialmente equivalenti, sia infine la 'Probit Analysis'.

Quest'ultima svolge l'analisi di regressione ponderata attraverso la trasformazione delle proporzioni di particolari categorie della variabile dipendente in unità normali standardizzate che generano le rispettive proporzioni come aree sottese alla curva normale (Finney, 1971).

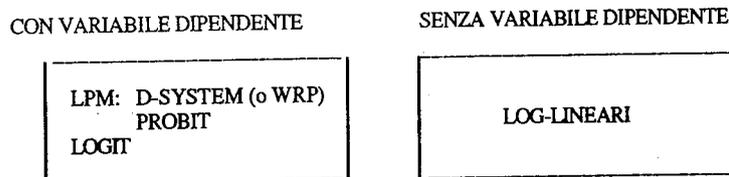
Alla classe dei modelli logit di cui abbiamo già parlato possono ricondursi le due versioni dei modelli che si trovano in letteratura e che assumono come variabile criterio i logaritmi degli odds o direttamente gli odds.

Infine esistono varie versioni di modelli log-lineari che non distinguono fra variabili dipendenti e indipendenti, a seconda che assumano come variabile criterio frequenze o log-frequenze, probabilità o log-probabilità.

Lo schema qui di seguito riportato riassume i vari elementi della classificazione dei modelli parametrici di analisi dei dati qualitativi.

Nelle pagine che seguono punteremo la nostra attenzione sull'analisi log-lineare e sulla analisi logit. La prima, in particolare, rappresenta il punto di partenza fondamentale per la comprensione delle altre tecniche parametriche di analisi dei dati qualitativi. Come è del resto nella tradizione della maggior parte dei ricercatori che si sono avvicinati a queste tecniche, più orientati alla survey research, utilizzeremo per i nostri esempi un problema di ricerca sostantivo relativo all'indagine longitudinale sulla selezione scolastica in provincia di Trento esposta in appendice.

Figura I: I modelli parametrici di analisi dei dati qualitativi



LPM: Linear Probability Models (M. Swafford, 1980)
D-SYSTEM: (J. Davis, 1976)
WRP: Weighted Regression for Proportions (J.E. Grizzle, 1969; D.G. Taylor, 1983)
PROBIT: normalized proportions (Finney, 1971)
LOGIT: log-odds models (L. Goodman, 1972; 1976)
LOG-LINEAR: log-frequencies o log-probability model (L. Goodman, 1972; S.J. Haberman, 1978)

Figura II: Applicazione dei modelli di regressione e di scomposizione lineare

		VARIABILE INDIPENDENTE		
		CONTINUA	DICOTOMICA	POLITOMICA
VARIABILE DIPENDENTE	CONTINUA	Regress. ordinaria (OLS)	Regress. Dummy (OLS)	ANOVA MCA
	DICOTOMICA	*	SCOMPOSIZIONE LINEARE PONDERATA (GSK) LPM (GLS) LOGIT (ML)	

Figura III: Caratteristiche e requisiti delle varie classi di modelli log-lineari (1)

CLASSI DI MODELLI	MODELLI SPECIFICI	VARIABILI CRITERIO	RANGE	FORMA E PARAMETRI	METODO DI STIMA		INTERPRETABILITA' DEI COEFFICIENTI STIMATI	
					N. GRANDE	N. PICCOLO	DIRETTA	INDIRETTA
LPM	D-SYSTEM (o WRP)	proporzioni	0/1	addit: $P = \beta_1 + \beta_2A + \beta_3B + \beta_4AB$	GLS	GSK	X	
	probit	proporz.-normal.	-3/+3	addit: $P' = \beta'_1 + \beta'_2A + \beta'_3B + \beta'_4AB$	GLS	GSK		X
LOGIT	logit	log-odds	1/log(N)	addit: $\log \Omega = \lambda_1 + \lambda_2A + \lambda_3B + \lambda_4AB$	ML	ML		X
	logit	odds	0/N	multipl: $\Omega = \gamma_1 \cdot \gamma_2^A \cdot \gamma_3^B \cdot \gamma_4^{AB}$	ML	ML	X	
LOG-LINEARI	log-lineare	frequenze	0/N	multipl: $F_{ij} = \eta \cdot \tau^A \cdot \tau^B \cdot \tau^{AB}$	ML	ML	X	
	log-lineare	log-frequenze	1/log(N)	addit: $\log(F) = \Theta + \lambda^A + \lambda^B + \lambda^{AB}$	ML	ML		X
	log-lineare	probab.	0/1	multipl: $p_{ij} = \mu \cdot \tau^A \cdot \tau^B \cdot \tau^{AB}$	ML	ML	X	
	log-lineare	log-probab.	$-\infty / 0$	addit: $\log(p) = \mu + \lambda^A + \lambda^B + \lambda^{AB}$	ML	ML		X

(1) Sono attualmente disponibili numerosi programmi di elaborazione automatica che eseguono l'analisi log-lineare

AUTORE E INDIRIZZO

Packages:

ECTA L. Goodman University of Chicago, 1126 East, 59th Street, Illinois 60637, USA
SPSSX Spss Inc., Suite 3300, 444 North Michigan Avenue, Chicago, Illinois, 60611
MAGHDSON J. Madigson, Abt Associates Inc., 55 Wheeler Street; Cambridge, MA 02138
C-TAB

International Educational Services, Box A3650, Chicago, Illinois, 60690

MULTIQUAL vedi: Analysis of qualitative data. Vol. 2: New developments, New York Academic Press
HABERMAN J. Richards Landis, Dept. of Biostatistics, The University of Michigan, Ann Arbor, Michigan, 48109
GENCAT Numerical Algorithm Group Ltd., 13 Banbury Road, Oxford, OX2 6NN, England
GLIM G.N. Gilbert, Depart. of Sociology, University of Surrey, Guildford, GU2 5XH, England
LOGLIN

CAPITOLO 1

Le varie forme di modelli log-lineari

1.1. - Il chi quadrato come misura della bontà di adattamento fra valori osservati e valori teorici.

Pur essendo il chi quadrato forse il test più conosciuto dai ricercatori sociali, lo riprendiamo brevemente in queste note in quanto la sua logica costituisce un elemento essenziale per la comprensione dei modelli log-lineari.

Dato che le frequenze attese provengono da precise ipotesi o modelli di distribuzione dei dati definiti dal ricercatore, è fondamentale disporre di un test che permetta la verifica statistica del confronto fra configurazioni osservate e modello ipotizzato.

Il test del chi quadrato, nella sua accezione di test di "goodness of fit", è la risposta più idonea a questo problema e trova una applicazione generale nella verifica dei modelli log-lineari.

Per fornire una prima esemplificazione di come il test operi proponiamo qui un esempio semplificato tratto dai dati della nostra ricerca sul comportamento scolastico¹.

Tabelle 1 - Proseguimento degli studi dopo la scuola dell'obbligo in provincia di Trento (a.s. 1973/74 - 1974/75)

	F _{ij}			P _{ij}		
	Maschi	Femmine	Totale	Maschi	Femmine	Totale
Continuano	2185	2012	4197	.339	.312	.651
Non continuano	1042	1212	2254	.161	.188	.354
Totale	3227	3224	6451	.500	.500	1.000

¹ Per le informazioni relative alla ricerca si veda l'appendice del volume.

Simboli di base adottati:

	F _{ij}			P _{ij}			e _{ij}		
	M	F	Tot.	M	F	Tot.	M	F	Tot.
C ₁	f ₁₁	f ₁₂	f _{1.}	p ₁₁	p ₁₂	p _{1.}	e ₁₁	e ₁₂	e _{1.}
C ₀	f ₂₁	f ₂₂	f _{2.}	p ₂₁	p ₂₂	p _{2.}	e ₂₁	e ₂₂	e _{2.}
Tot.	f _{.1}	f _{.2}	f _{..}	p _{.1}	p _{.2}	p _{..}	e _{.1}	e _{.2}	e _{..}

$$f_{i.} = \sum_j f_{ij}$$

$$p_{i.} = \sum_j p_{ij}$$

$$e_{i.} = \sum_j e_{ij}$$

$$f_{.j} = \sum_i f_{ij}$$

$$p_{.j} = \sum_i p_{ij}$$

$$e_{.j} = \sum_i e_{ij}$$

$$f_{..} = \sum_i \sum_j f_{ij}$$

$$p_{..} = \sum_i \sum_j p_{ij}$$

$$e_{..} = \sum_i \sum_j e_{ij}$$

Da una prima lettura dei dati appare improponibile il modello di *equiprobabilità*, indicato di regola con (), che assume un'identica distribuzione dei soggetti all'interno delle celle (CONTINUANO/NON CONTINUANO) della tavola di contingenza. Tale ipotesi non terrebbe in alcuna considerazione i vincoli rappresentati dai marginali di riga e di colonna, vale a dire della diversa distribuzione fra coloro che continuano gli studi e coloro che non continuano nè delle differenze fra i maschi e le femmine.

Tale ipotesi prevederebbe la seguente configurazione:

	(e _{ij})			(p _{ij})		
	Maschi	Femmine	Totale	Maschi	Femmine	Totale
C ₁	1612.75	1612.75	3225.5	0.25	0.25	0.50
C ₀	1612.75	1612.75	3225.5	0.25	0.25	0.50
Tot.	3225.50	3225.50	6451.0	0.50	0.50	1.00

Una seconda ipotesi da sottoporre a test è il modello di *indipendenza* fra proseguimento degli studi e sesso dell'allievo, d'ora in poi indicato con (C)(S), secondo il quale la distribuzione dei soggetti all'interno delle celle è rigidamente casuale con l'unico vincolo rappresentato dai marginali di riga e di colonna. Le ipotesi di indipendenza avrebbero le seguenti configurazioni:

Indipendenza di S: (C)

	Maschi	Femmine	Totale
Continuano	2098.5	2098.5	4197
Non continuano	1127.0	1127.0	2254
Totale	3225.5	3225.5	6451

Indipendenza di C: (S)

	Maschi	Femmine	Totale
Continuano	1613.5	1612	3225.5
Non continuano	1613.5	1612	3225.5
Totale	3227.0	3224	6451.0

Indipendenza di S e C: (C)(S)

	Maschi	Femmine	Totale
Continuano	2099.5	2097.5	4197
Non continuano	1127.5	1126.5	2254
Totale	3227.0	3224.0	6451

Secondo il teorema della probabilità composta applicato al nostro esempio, la probabilità di continuare gli studi per un allievo maschio è pari al prodotto della probabilità di continuare per la probabilità di essere maschio.

Nei noti simboli:

$p_{ij} = p_i \cdot p_j$ da cui, utilizzando le frequenze,

$$p_{ij} = \frac{f_{i.}}{f_{..}} \cdot \frac{f_{.j}}{f_{..}} = \frac{f_{i.} \cdot f_{.j}}{f_{..}^2}$$

$$e_{ij} = p_{ij} \cdot f_{..} = \frac{f_{i.} \cdot f_{.j}}{f_{..}} \quad \text{da cui} \quad e_{ij} = \frac{f_{i.} \cdot f_{.j}}{f_{..}}$$

Questa formula è la nota espressione abbreviata del prodotto dei marginali utilizzato per ottenere le varie frequenze attese ai fini del calcolo del chi quadrato.

L'applicazione della formula produce la serie delle frequenze attese e_{ij} nel modello di indipendenza da confrontare con le frequenze osservate F_{ij} , come sintetizzato in Tabella 2.

Tabella 2 - Calcolo del chi quadrato secondo il modello di indipendenza.

	F_{ij}	e_{ij}	$(F_{ij} - e_{ij})$	$(F_{ij} - e_{ij})^2$	$\frac{(F_{ij} - e_{ij})^2}{e_{ij}}$
Continuano/maschi	2185	2099.5	85.5	7310.25	3.482
Continuano/femmine	2012	2097.5	-85.5	7310.25	3.482
Non continuano/maschi	1042	1127.5	85.5	7310.25	6.484
Non continuano/femmine	1212	1126.5	-85.5	7310.25	6.489
				Totale	19.940

Chi quadrato di Pearson per la bontà di adattamento:

$$\chi_1^2 = \sum_{ij} \left(\frac{F_{ij} - e_{ij}}{e_{ij}} \right)^2 = 19.940$$

Il valore χ^2 trovato risulta significativo allo 0.01 dal confronto con la tavola dei valori teorici con un grado di libertà.

L'ipotesi di indipendenza tra sesso e continuazione degli studi è pertanto da rifiutare e quindi il "modello di indipendenza" non si adatta bene ai dati osservati.

Un test alternativo di larga applicazione nell'analisi log-lineare è il test "Rapporto di verosimiglianza" (Likelihood ratio).

$$L = \prod_{ij} \left(\frac{e_{ij}}{F_{ij}} \right)^{F_{ij}}$$

Su questa espressione si fonda una seconda misura di chi quadrato denominata "Chi quadrato rapporto di verosimiglianza" (Likelihood ratio chi square) la cui formula è:

$$L^2 = 2 \sum_{ij} F_{ij} \text{Log} \left(\frac{e_{ij}}{F_{ij}} \right)$$

Il vantaggio fondamentale del test L^2 è la sua scomponibilità. Esso consente di confrontare direttamente fra loro diversi modelli teorici che presuppongono ipotesi più o meno restrittive e per questo motivo nel

confronto fra modelli gerarchici verrà utilizzato prevalentemente il test L^2 .²

L'ipotesi che la distribuzione dei valori interni di una tabella di contingenza rispetti le proporzioni dei marginali consente una notevole riduzione del numero di parametri da stimare. Infatti, dovendo stimare le singole probabilità di cella p_{ij} di una tabella 2×2 , possiamo procedere in diversi modi secondo ipotesi progressivamente strutturate:

a) Innanzitutto si può partire dai singoli valori osservati ipotizzando che ciascuna probabilità sia vincolata soltanto dalla numerosità del campione, vale a dire:

$$p_{ij} = \frac{f_{ij}}{f_{..}}$$

In tale caso occorre calcolare tre valori su quattro per ottenere le probabilità corrispondenti. Si hanno 4 dati, 1 vincolo rappresentato dalla numerosità campionaria e quindi 3 gradi di libertà.

Nel caso, analogo, di tabelle a tre dimensioni il numero dei valori da calcolare, sotto l'unico vincolo dell' N totale, cresce rapidamente. Ad esempio, nel caso di tabelle $2 \times 2 \times 2$, si hanno 7 valori da calcolare; nel caso di tabelle $4 \times 4 \times 2$ si hanno 31 valori, ecc..

b) Se viceversa ipotizziamo il vincolo del rispetto delle frequenze marginali (ipotesi di indipendenza)

$$p_{ij} = P_i \cdot P_j$$

² Un ulteriore motivo di preferenza per l' L^2 sta nel fatto che il Chi quadrato talvolta presenta valori negativi quando viene testata la significatività di effetti individuali sul modeling gerarchico, valori che devono essere interpretati come zero (Taylor, 1983, p. 561).

i parametri da stimare si riducono ad uno, avendo introdotto, oltre al vincolo del rispetto delle frequenze totali, i due vincoli di rispetto delle due marginali (nell'esempio, la continuazione degli studi ed il sesso).

Secondo questa ipotesi di indipendenza, in una tabella 2×2 occorre calcolare un solo valore per ottenere le probabilità (o le frequenze attese) corrispondenti. Si hanno infatti 4 dati, 3 vincoli ed un grado di libertà. Nel caso di una tabella $4 \times 4 \times 2$ il vantaggio è più vistoso: anziché 31 parametri se ne devono calcolare soltanto sette ($3+3+1$), sempre secondo l'ipotesi di indipendenza.

$$P_{ijk} = P_{i..} \cdot P_{.j.} \cdot P_{..k}$$

c) Se i vincoli ipotizzati nel modello aumentano e comprendono l'interazione fra due o più variabili della tabella di contingenza, i gradi di libertà si riducono ulteriormente fino ad arrivare a zero qualora tutte le interazioni di primo, secondo ed ulteriore grado siano state incluse.

Così, ad esempio, qualora si ipotizzi un'interazione fra continuazione degli studi e sesso, nel senso che la probabilità di continuazione degli studi sia diversa per maschi e femmine (67,7% M; 62,4% F), l'ulteriore vincolo inserito azzeri i gradi di libertà e le frequenze attese, secondo questa ipotesi di interazione, finiscono per coincidere con quelle osservate. Un modello di questo tipo è detto modello SATURO che indicheremo, come è d'uso, con (CS).

In tabelle a tre o più dimensioni si possono inserire come vincoli ipotesi di interazione di primo e secondo ordine che riducono i gradi di libertà.

In una tabella 3×3 , ad esempio, si possono inserire nel modello fino a 4 ipotesi di interazione di primo ordine, definite nell'ambito delle singole celle della tabella stessa tenendo conto del fatto che i gradi di libertà rimasti dopo i due precedenti vincoli (del totale generale e dei marginali) sono quattro.

Tabelle 3 - Calcolo dei gradi di libertà

Tabelle [2x2]

	Modelli				
	()	(C)	(S)	(C)(S)	(CS)
N. celle	4	4	4	4	4
Vincoli	1	2	2	3	4
G.L.	3	2	2	3-1	0

Tabelle [3x3]

	()	(T)	(C)	(C)(T)	(C _i T _j)			
N. celle	9	9	9	9	9	9	9	9
Vincoli	1	3	3	5	6	7	8	9
G.L.	8	6	6	4	3	2	1	0

Tabelle [rxc]

	()	(T)	(C)	(C)(T)	(C _i T _j)			
N.celle	rc	rc	rc	rc	rc	rc	rc	rc
Vincoli	1	(r-1)+1	(c-1)+1	(r-1)(c-1)+1	(r-1)(c-1)	(r-1)(c-1)-1	(r-1)(c-1)-2	(r-1)(c-1)-3
g.l.	(rc-1)	(rc-r)	(rc-c)	r+c-2	r+c-1	r+c	r+c+1	rc

Riprendendo il nostro esempio, passiamo ora alla costruzione del modello saturo che include l'interazione fra continuazione degli studi (C) e sesso (S).

Tabelle 4 - Frequenze osservate ed attese nel modello saturo (CS) 2x2

		Maschi	Femmine	Totale
Continuano	f _{ij}	2185	2012	4197
	e _{ij}	2185	2012	4197
Non continuano	f _{ij}	1042	1212	2154
	e _{ij}	1042	1212	2154
Totali	f _{..}	3227	3224	
	e _{..}	3227	3224	

Nelle tabelle a due dimensioni l'inserimento di ipotesi di interazione non porta necessariamente alla saturazione del modello. Mentre ciò avviene sempre nelle tabelle 2x2, che hanno un solo vincolo per l'interazione, nel caso di variabili politomiche, l'interazione può avvenire a diversi livelli e combinazioni tra le due variabili e come tale essere inclusa fra le ipotesi.

Ad esempio, in una tabella 4x2 (continuazione degli studi per sesso), in cui la continuazione degli studi viene articolata su quattro livelli:

C₁ = continuano serie A (licei, ecc.);

C₂ = continuano serie B (ist. tecnici, ecc.);

C₃ = continuano serie C (ist. professionali, ecc.);

C₄ = non continuano;

si potranno definire fino a tre ipotesi di interazione fra tipo di studi scelto e sesso.

Tabella 5 - Frequenze osservate e attese nei vari modelli 4x2 e calcolo dei gradi di libertà.

A) Frequenze osservate [f _{ij}]				B) Modello di indipendenza [e _{ij}] (C) (S)			
	M	F	Totale		M	F	Totale
C ₁	334	278	612	C ₁	304	308	612
C ₂	839	685	1524	C ₂	758	766	1524
C ₃	897	971	1868	C ₃	929	939	1868
C ₄	911	1079	1990	C ₄	990	1000	1990
Tot.	2981	3013	5994	Tot.	2981	3013	5994

$$\chi^2_3 = 38.04 (p < 0.01)$$

C) Modelli con interazione (1)

(C ₂ S)		(C ₃ S)		(C ₄ S)	
M	F	M	F	M	F
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
839	685	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	897	971	<input type="checkbox"/>	<input type="checkbox"/>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	911	979
g.l.=2		g.l.=2		g.l.=2	
(C ₂ S)	(C ₃ S)	(C ₂ S)	(C ₄ S)	(C ₃ S)	(C ₄ S)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
839	685	839	685	<input type="checkbox"/>	<input type="checkbox"/>
897	971	<input type="checkbox"/>	<input type="checkbox"/>	897	971
<input type="checkbox"/>	<input type="checkbox"/>	911	1079	911	1079
g.l.=1		g.l.=1		g.l.=1	
(C ₂ S)	(C ₃ S)(C ₄ S) ecc.				
<input type="checkbox"/>	<input type="checkbox"/>				
839	685				
897	971				
911	1079				
g.l. = 0					

(1) I livelli equivalgono qui all'ordine dell'interazione considerando che la variabile polinomica C a 4 livelli equivale a 3 variabili Dummy.

Tabelle 4x2

	Dati	Vincoli	g.l.
Equiprobabilità	8	1	7
Indipendenza	8	5	3
Interazione: 1 livello	8	6	2
2 livello	8	7	1
3 livello	8	8	0

Dai dati riportati nella tabella delle frequenze osservate si può ragionevolmente ipotizzare un'interazione fra i livelli C₂ e C₄ e il sesso. Fra le varie ipotesi esemplificate nella tabella n. 5 C), l'interazione (C₂S) (C₄S) infatti è quella che produce una maggiore riduzione del chi quadrato:

Nella tabella sono esemplificati solo alcuni dei possibili modelli con interazione su 1, 2 o 3 livelli della variabile "continuazione degli studi" lasciando in evidenza come le celle non vincolate dall'ipotesi di interazione rimangono libere di fluttuare entro i vincoli assegnati dai totali marginali di riga e di colonna. L'assegnazione dei valori in tali celle è di difficile realizzazione manuale per cui è necessario ricorrere ad appositi algoritmi iterativi di calcolo (Iterative proportional fitting). Il più conosciuto di questi algoritmi è quello di Stephan-Deming, utilizzato da L. Goodman e descritto in numerosi testi (Davis, 1974; Fienberg, 1977; Haberman, 1979; Bishop, Fienberg, Holland, 1975) ai quali rinviamo per i dettagli matematici.

L'algoritmo di Stephan-Deming è un metodo che serve per generare delle frequenze attese nelle celle di una tavola di contingenza assumendo una certa struttura nelle frequenze della tabella stessa dove la somma per riga e per colonna dei dati osservati non coincide con i marginali teorici. Il risultato dell'algoritmo è quindi una tabella nella quale le frequenze di cella hanno la stessa struttura della tabella di partenza ma i marginali sono identici ai marginali teorici. Per 'struttura' si intendono rapporti fra prodotti delle frequenze di certe celle; le celle utilizzate per formare i prodotti (e perciò i rapporti) dipendono dalla struttura che viene definita. Quanto al

funzionamento dell'algoritmo, esso procede in maniera iterativa e per passi successivi; prima vengono soddisfatti i vincoli di riga dividendo le relative frequenze per il corrispondente marginale osservato e moltiplicandolo per quello teorico. Le frequenze così ottenute vengono successivamente aggiustate con lo stesso procedimento per soddisfare i vincoli delle colonne. Poichè questa operazione comporta l'alterazione dei valori ottenuti con il primo passo e quindi i vincoli di riga non sono più soddisfatti, si ripete la procedura fino a quando le differenze fra valore teorico e valore osservato divengono inferiori ad un valore q (criterio di convergenza) prefissato (Stephan F.F., Deming W.E., 1940; Corbetta P. Schadee, 1984).

1.2. - Modelli log-lineari per tabelle 2x2

Prima di formulare un modello che descriva tutte le complesse interrelazioni di una tabella multidimensionale, consideriamo la situazione più semplice costituita dalle tabelle 2x2 utilizzando l'esempio della continuazione degli studi (C) per sesso (S) in provincia di Trento.

Tabella 6 - Proseguimento degli studi dopo la scuola dell'obbligo in provincia di Trento per sesso

	F _{ij}			P _{ij}		
	Maschi	Femmine	Totale	Maschi	Femmine	Totale
Continuano C ₁	2185	2012	4197	.339	.312	.651
Non continuano C ₂	1042	1212	2254	.161	.188	.349
Totale	3227	3224	6451	.500	.500	1000

Differenze percentuali e chi quadrato.

Il modo tradizionale per identificare una relazione fra due variabili categoriali è quello di calcolare le percentuali (continuano/non continuano) entro le categorie della variabile dipendente (Maschi/Femmine) e confrontarle fra loro. Se le percentuali differiscono per un ammontare significativo (usando il normale test del chi quadrato per indipendenza) si può affermare l'esistenza della relazione.

Una prima lettura della tabella permette di formulare l'ipotesi di indipendenza, testando le differenze tra le percentuali con l'usuale test chi quadrato ($\chi^2 = 19,94$). Le percentuali di maschi e femmine che continuano gli studi sono le seguenti:

	Maschi	Femmine
Continuano	67.7	62.4
Non continuano	32.3	37.6
Totale	100.0	100.0

Esiste pertanto una relazione tra sesso e continuazione degli studi: *in provincia di Trento i maschi continuano gli studi più delle femmine.*

I rapporti di probabilità (odds).

L'esistenza di una relazione fra sesso e proseguimento degli studi può venire evidenziata attraverso una procedura messa a punto da Davis (1976), che calcola i rapporti di probabilità (odds) nella tabella.

Al fine di utilizzare i modelli log-lineari si rende necessaria una riconcettualizzazione della variabile dipendente (D. Knoke - P.J. Burke, 1980, p. 9). Al posto delle proporzioni, in cui le frequenze di cella vengono divise per il totale marginale, la forma base della variazione da spiegare è costituita dagli odds. "Un odds è il rapporto fra la frequenza di coloro che sono in una categoria e la frequenza di coloro che non sono in

quella categoria. La sua interpretazione è la possibilità che un individuo scelto a caso si trovi in una data categoria prestabilita piuttosto che in una qualunque delle altre" (Knoke - Burke, 1980, p. 9).

Nel nostro esempio della tabella 6 l'odds per un soggetto di proseguire o non proseguire negli studi è dato da 4197/2254 vale a dire un odds di circa 2 contro 1. Tale odds è un ODDS MARGINALE, calcolato sulle frequenze totali di una categoria indipendentemente dalle categorie dell'altra variabile. Possiamo calcolare anche gli ODDS CONDIZIONATI nel corpo della tabella che corrispondono alle percentuali tradizionali. Essi rappresentano le probabilità di continuare gli studi rapportate alla probabilità di non continuare dato il particolare livello della variabile sesso. In altre parole gli odds condizionati sono rapporti di probabilità fra i parziali dell'altra variabile.

Nel nostro esempio gli odds condizionati riferiti alla variabile continuazione degli studi sono i seguenti:

$$\text{Per i maschi: } \frac{\text{continuano}}{\text{non continuano}} = \frac{2185}{1042} = 2.097$$

$$\text{Per le femmine: } \frac{\text{continuano}}{\text{non continuano}} = \frac{2012}{1212} = 1.660$$

Così come l'inesistenza di una relazione implica un ammontare uguale nelle percentuali di continuazione degli studi per maschi e femmine, così, nell'ottica degli odds, tale situazione è evidenziata dall'uguaglianza degli odds condizionati. Il rapporto tra gli odds condizionati, detto ODDS RATIO, fornisce la misura dell'esistenza della relazione.

Nel nostro esempio:

$$\text{ODDS RATIO} = \frac{2.097}{1.660} = 1.263$$

L'esistenza di una associazione fra sesso e continuazione degli studi è evidenziata appunto da un odds ratio (rapporto fra i rapporti di probabilità fra continuazione/non continuazione per i maschi e per le femmine) maggiore di 1.

Il rapporto continuano/non continuano è pertanto maggiore per i maschi i quali tendono a proseguire gli studi dopo la terza media in provincia di Trento in misura superiore delle femmine.

Come fanno giustamente rilevare Knoke e Burke (1980, p. 10) "gli odds-ratio sono il concetto base dei modelli log-lineari". Vale pertanto la pena di analizzarne ulteriormente la struttura e le interpretazioni.

Tabella 7 - Calcolo degli ODDS nella tabella n. 1

ODDS MARGINALI:

$$\text{ODDS(C)} = \frac{4197}{2254} = 1.862$$

$$\text{ODDS(S)} = \frac{3227}{3224} = 1.001$$

ODDS CONDIZIONATI:

$$\text{Per C: } \text{ODDS (C/M)} = \frac{F_{11}}{F_{12}} = \frac{2185}{1042} = 2.097$$

$$\text{ODDS (C/F)} = \frac{F_{12}}{F_{22}} = \frac{2012}{1212} = 1.660$$

$$\text{Per S: } \text{ODDS (S/C}_1) = \frac{F_{11}}{F_{12}} = \frac{2185}{2012} = 1.086$$

$$\text{ODDS (S/C}_2) = \frac{F_{21}}{F_{22}} = \frac{1042}{1212} = 0.860$$

(segue)

Tabella 7: (segue)

$$\text{ODDS RATIO: } \frac{f_{11}}{f_{21}} \cdot \frac{f_{12}}{f_{22}} = \frac{2.097}{1.660} = 1.263$$

$$\text{oppure (prodotto incrociato)} \frac{f_{11} f_{22}}{f_{21} f_{12}} = \frac{2185 \cdot 1212}{1042 \cdot 2012} = 1.263$$

$$\text{oppure } \frac{f_{11}}{f_{12}} \cdot \frac{f_{21}}{f_{22}} = \frac{1.086}{0.860} = 1.263$$

$$\text{Q di Yule: } \frac{\text{odds ratio} - 1}{\text{odds ratio} + 1} = \frac{f_{11} f_{22} - f_{12} f_{21}}{f_{11} f_{22} + f_{12} f_{21}}$$

$$\frac{1.263 - 1}{1.263 + 1} = \frac{2185 \cdot 1212 - 2012 \cdot 1042}{2185 \cdot 1212 + 2012 \cdot 1042}$$

$$\frac{0.263}{2.263} = \frac{551.716}{4.744.724} = 0.116$$

$$\text{Significatività di Q: } Q_{CS} \pm Z_C \sqrt{\frac{(1-Q_{CS}^2)^2 \left(\frac{1}{f_{11}} + \frac{1}{f_{12}} + \frac{1}{f_{21}} + \frac{1}{f_{22}} \right)}{4}}$$

per $Z_C = 2.58$ ($p < 0.01$)

$$0.116 \pm 2.58 \sqrt{\frac{1}{4} (1-0.116^2)^2 \left(\frac{1}{2185} + \frac{1}{2012} + \frac{1}{1042} + \frac{1}{1212} \right)}$$

$$0.116 \pm 2.58 \cdot 0.0258$$

$$0.116 \pm 0.0666$$

$$0.049 < Q_{CS} < 0.18$$

2^a Risposta: Q_{CS} è significativo allo 0.01. Esiste una relazione fra continuazione degli studi e sesso.

Data la seguente tabella di frequenze osservate (F_{ij})

	F_{ij}		
	S_1	S_2	
C_1	f_{11}	f_{12}	$f_{1.}$
C_2	f_{21}	f_{22}	$f_{2.}$
	$f_{.1}$	$f_{.2}$	$f_{..}$

l'odds ratio può esprimersi così:

$$\text{ODDS RATIO (CS) osservato} = \frac{f_{11}}{f_{21}} \cdot \frac{f_{12}}{f_{22}}$$

che, semplificato, diviene il noto "prodotto in croce" per tabelle 2x2:

$$\text{ODDS RATIO (CS)} = \frac{f_{11} f_{22}}{f_{21} f_{12}}$$

E' interessante notare come il Q di Yule, classica misura di associazione per tabelle 2x2, sia una semplice funzione dell'odds ratio:

$$Q = \frac{\text{odds ratio} - 1}{\text{odds ratio} + 1} = \frac{(f_{11}) (f_{22}) - (f_{12}) (f_{21})}{(f_{11}) (f_{22}) + (f_{12}) (f_{21})}$$

Mentre il Q di Yule ha un campo di variabilità compreso fra -1 e +1, gli odds ratio assumono soltanto valori positivi, non hanno limite superiore ed hanno valore 1.00 quando non vi è alcuna relazione (i due odds condizionati sono uguali).

Odds ratio superiori ad 1 indicano diretta covarianza tra variabili mentre odds ratio inferiori ad 1 indicano una relazione inversa.

Naturalmente la "direzione" della covariazione è arbitraria quando le variabili sono misurate a livello di scala nominale dato che l'ordine delle categorie può essere mutato. Nel nostro esempio, come indicato nella Tabella 6, l'odds ratio ammonta a 0.860 qualora il calcolo venga fatto sugli odds condizionati calcolati per sesso.

Modelli log-lineari

Nè la procedura delle percentuali nè quella degli odds permettono una completa quantificazione diretta dei tre effetti (riga, colonna, interazione). Si rende pertanto necessaria la definizione di un metodo matematico che consenta di quantificare l'importanza relativa degli effetti che si riscontrano in una tabella di contingenza e di stabilire se essi siano significativi. Tale modello, suggerito per la prima volta da Leo Goodman (1972) si fonda, nella sua versione originaria, sui logaritmi naturali delle probabilità di cella (p_{ij}).

Esistono quattro versioni principali del modello log-lineare di Goodman che passeremo in rassegna in questo capitolo.

Una prima versione opera con le log-probabilità di cella ($v_{ij} = \log p_{ij}$) in un modello di tipo additivo.

La sua forma generale è:

$$v_{ij} = \mu + \lambda^C + \lambda^S + \lambda^{CS}$$

La seconda versione opera, sempre in forma additiva, sui logaritmi delle frequenze di cella (o log-frequenze) ($y_{ij} = \log F_{ij}$).

La sua forma generale (per modelli saturi) è:

$$y_{ij} = \Theta + \lambda^C + \lambda^S + \lambda^{CS}$$

La terza versione lavora in forma moltiplicativa sulle probabilità di cella (p_{ij}).

La sua forma generale è:

$$p_{ij} = u \tau^C \tau^S \tau^{CS}$$

La quarta versione opera, sempre in forma moltiplicativa, sulle frequenze di cella (F_{ij}).

La sua forma generale è:

$$F_{ij} = \eta \tau^C \tau^S \tau^{CS}$$

Malgrado le differenze formali dei quattro modelli, il calcolo dei parametri relativi agli effetti riga, colonna e interazione è riconducibile agli odds ratio appena ricordati. La classe dei modelli additivi deriva i λ^{CS} (sia che operi sulle log-probabilità che sulle log-frequenze: i valori dei parametri λ sono gli stessi nei due modelli!) dai logaritmi degli odds ratio³

Vale la seguente relazione:

$$\lambda^{CS} = \frac{\text{Log odds}}{\text{N. celle}}$$

Il valore del parametro λ^{CS} indica la componente interazione del modello additivo ed è pari al valore medio (geometrico) dell'odds ratio di cella.

³ Dato che le probabilità si stimano con le frequenze, i due modelli I e II della tabella 8 differiscono soltanto per quanto riguarda l'intercetta (μ e Θ rispettivamente), mentre i valori λ sono uguali.

Tabella 8 - Le quattro classi di modelli log-lineari

	Additivo	Moltiplicativo
Probabilità	$v_{ij} = \log p_{ij}$ $v_{ij} = \mu + \lambda^C + \lambda^S + \lambda^{CS}$	P_{ij} $p_{ij} = u \tau^C \tau^S \tau^{CS}$
Frequenze	$y_{ij} = \log F_{ij}$ $y_{ij} = \Theta + \lambda^C + \lambda^S + \lambda^{CS}$	F_{ij} $F_{ij} = \eta \tau^C \tau^S \tau^{CS}$

La relazione tra odds ratio e parametro λ^{CS} può essere colta agevolmente nella seguente sequenza:

$$\begin{array}{c}
 \text{ODDS RATIO} \\
 \downarrow \\
 \log (\text{ODDS RATIO}) \\
 \downarrow \\
 \log \left(\frac{\text{ODDS RATIO}}{N. \text{ celle}} \right) = \lambda^{CS} \text{ (odds ratio medio di cella)}
 \end{array}$$

La classe dei modelli moltiplicativi deriva i τ^{CS} (anche questi sono identici sia che si usino probabilità o frequenze) direttamente dagli odds ratio.

Vale infatti la seguente relazione:

$$\tau^{CS} = (\text{odds ratio})^{1/n.\text{celle}}$$

Anche in questo caso il valore del parametro τ^{CS} indica la componente interazione del modello moltiplicativo ed è pari al valore medio dell'odds

ratio di cella.

La sequenza è in questo caso più rapida:

$$\begin{array}{c}
 \text{ODDS RATIO} \\
 \downarrow \\
 (\text{ODDS RATIO})^{1/r \ c} = \tau^{CS} \text{ (odds ratio medio di cella)}
 \end{array}$$

Riprendendo i dati della nostra tabella 2x2, relativa alla continuazione degli studi per sesso, abbiamo:

$$\lambda^{CS} = \frac{\log(\text{odds ratio})}{n. \text{ celle}} = \frac{\log 1.263}{4} = \frac{0.233}{4} = 0.058$$

$$\tau^{CS} = (\text{odds ratio})^{1/r \ c} = (1.263)^{1/4} = 1.060$$

Come vedremo, i modelli additivi e moltiplicativi si corrispondono in modo perfetto essendo due modi che, pur diversi formalmente, esprimono le stesse grandezze (frequenze di cella).

Si ha infatti:

$$\lambda^{CS} = \log \tau^{CS}$$

sostituendo i termini con i dati della tabella abbiamo:

$$\lambda^{CS} = \log 1.060 = 0.058$$

1.2.1. - Forma moltiplicativa del modello

Per una più completa analisi del modello log-lineare vediamo ora il modello saturo nella versione moltiplicativa di Goodman per una tabella 2x2⁴. Il termine "saturo" significa che sono presenti nel modello tutti i possibili effetti parametro.

Tale modello presenta la forma:

$$F_{ij} = \eta \tau_i^C \tau_j^S \tau_{ij}^{CS} \quad [1]$$

dove F_{ij} rappresenta l'ammontare delle frequenze attese della cella ij ; η (η) è la media geometrica del numero di casi in ciascuna cella della tabella ed è un termine molto simile all'intercetta dell'equazione di regressione, una specie di "fattore di scala" della tabella. Essa rappresenta un punto di partenza per la misura degli effetti pur non avendo un preciso significato sostantivo. I termini τ (tau) rappresentano gli effetti che le variabili hanno sulle frequenze di cella legati, come si è visto, agli odds e agli odds ratio. Gli effetti τ_i^C (uno per ciascun livello di C) sono presenti quando esiste una distribuzione asimmetrica (non rettangolare) nei marginali nella variabile "continuazione degli studi". Gli effetti τ_{ij}^S , analogamente, rappresentano il grado di asimmetria della variabile sesso.

Infine, gli effetti τ_{ij}^{CS} (uno per ciascuna delle celle ij della tavola) rappresentano l'interazione fra continuazione e sesso. Dati questi nove parametri effetto, le quattro frequenze di cella attese della Tabella 6 possono essere rappresentate dal modello come il prodotto di una serie di termini così come risulta dalla Tabella n. 9.

⁴ Per questa rassegna ci riferiamo fedelmente al fondamentale contributo di D. Knoke, P.J. Burke (1980, pp. 11-29).

Tabella 9 - Parametri e vincoli nei modelli log-lineari per tabelle 2x2

A) Modelli moltiplicativi

Equazione modello saturo: $F_{ij} = \eta \tau_i^C \tau_j^S \tau_{ij}^{CS}$

9 parametri:

	[S]		Totale
	S1	S2	
C1	τ_{11}^{CS}	τ_{12}^{CS}	τ_1^C
[C]			
C2	τ_{21}^{CS}	τ_{22}^{CS}	τ_2^C
Totale	τ_1^S	τ_2^S	η

5 vincoli: $\tau_{11}^{CS} = \tau_{22}^{CS} = \frac{1}{\tau_{12}^{CS}} = \frac{1}{\tau_{21}^{CS}}; \tau_1^C = \frac{1}{\tau_2^C}; \tau_1^S = \frac{1}{\tau_2^S}$

4 incognite - frequenze attese di cella da stimare: $f_{11}, f_{12}, f_{21}, f_{22}$

B) Modelli additivi

Equazione modello saturo: $\log F_{ij} = y_{ij} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS}$

9 parametri:

	[S]		Totale
	S1	S2	
C1	λ_{11}^{CS}	λ_{12}^{CS}	λ_1^C
[C]			
C2	λ_{21}^{CS}	λ_{22}^{CS}	λ_2^C
Totale	λ_1^S	λ_2^S	Θ

5 vincoli: $\lambda_2^C = -\lambda_1^C; \lambda_2^S = -\lambda_1^S; \lambda_{22}^{CS} = -\lambda_{12}^{CS} = -\lambda_{21}^{CS} = \lambda_{11}^{CS}$

4 incognite - frequenze attese di cella da stimare: $f_{11}, f_{12}, f_{21}, f_{22}$

Tabella 10 - Versione moltiplicativa di Goodman

(Modello saturo: $F_{ij} = \eta \tau_i^C \tau_j^S \tau_{ij}^{CS}$)

		S(sesto)	
		MASCHI	FEMMINE
C	C ₁ Continuano	$F_{11} = \eta \tau_1^C \tau_1^S \tau_{11}^{CS}$ 2185	$F_{12} = \eta \tau_1^C \tau_2^S \tau_{12}^{CS}$ 2012
	C ₂ Non continuano	$F_{21} = \eta \tau_2^C \tau_1^S \tau_{21}^{CS}$ 1042	$F_{22} = \eta \tau_2^C \tau_2^S \tau_{22}^{CS}$ 1212

Modello frequenze attese

$$\eta = (F_{11} F_{12} F_{21} F_{22})^{1/4} = (2185 \cdot 2012 \cdot 1042 \cdot 1212)^{1/4} = 1535.0$$

(Media geometrica delle frequenze F_{ij})

odds products 2.097 1.660

$$\tau^C = \left(\frac{F_{11}}{F_{21}} \cdot \frac{F_{12}}{F_{22}} \right)^{1/4}$$

$$\tau^C = \left(\frac{2185}{1042} \cdot \frac{2012}{1212} \right)^{1/4} = 3.4818^{1/4} = 1.366$$

marginal odds (C) medio di cella

$$\tau^S = \left(\frac{F_{11}}{F_{12}} \cdot \frac{F_{21}}{F_{22}} \right)^{1/4}$$

$$\tau^S = (1.086 \cdot 0.860)^{1/4} = 0.983$$

marginal odds (S) medio di cella

$$\tau^{CS} = \left(\frac{F_{11}}{F_{21}} \cdot \frac{F_{12}}{F_{22}} \right)^{1/4}$$

$$\tau^{CS} = \left(\frac{2.097}{1.660} \right)^{1/4} = (1.263)^{1/4} = 1.060$$

$$\tau^{CS} = \begin{cases} \text{effetto CS medio per cella} \\ \text{odds ratio medio di cella} \end{cases}$$

(segue)

Tabella 10 : (segue)

Stime: formule operative

$$\hat{\eta} = 1.535$$

$$\tau_1^C = \frac{1}{\tau_2^C} = \frac{(f_{11} f_{12})^{1/2}}{\eta} = \frac{(2185 \cdot 2012)^{1/2}}{1535} = \frac{2096,7}{1535} = 1.366$$

$$\tau_1^S = \frac{1}{\tau_2^S} = \frac{(f_{11} f_{21})^{1/2}}{\eta} = \frac{(2185 \cdot 1042)^{1/2}}{\eta} = 0.983$$

$$\tau_{11}^{CS} = \tau_{22}^{CS} = \frac{1}{\tau_{12}^{CS}} = \frac{1}{\tau_{21}^{CS}} = \left(\frac{f_{22}}{f_{21}} \cdot \frac{f_{22}}{f_{12}} \right)^{1/4} = \left(\frac{2185 \cdot 1212}{1042 \cdot 2012} \right)^{1/4} = \left(\frac{2.097}{1.660} \right)^{1/4} = (1.263)^{1/4} = 1.060$$

	η	τ_i^C	τ_j^S	τ_{ij}^{CS}	F_i
$F_{11} =$	1.535	· 1.366	· 0.983	· 1.060	$= 2.185$
$F_{12} =$	1.535	· 1.366	· 1.017	· 0.943	$= 2.012$
$F_{21} =$	1.535	· 0.732	· 0.983	· 0.943	$= 1.042$
$F_{22} =$	1.535	· 0.732	· 1.017	· 1.060	$= 1.212$

A parte il termine η , che rappresenta una frequenza media di cella, la grandezza di un effetto è misurata come scostamento dal valore di 1.00, valore equivalente ad un effetto nullo che lascia il prodotto inalterato. Se non vi fosse alcun effetto (ipotesi di equiprobabilità), le frequenze di cella sarebbero tutte uguali fra loro e uguali al valore del termine η .

Quando i parametri sono superiori a 1.00 le frequenze osservate saranno superiori alle frequenze attese; viceversa quando essi sono inferiori a 1.00.

Per variabili dicotomiche, come la continuazione degli studi e il sesso, i parametri τ per la categoria di ciascuna variabile sono reciproci e cioè valgono i seguenti vincoli ⁵ :

$$\tau^C = \tau_1^C = 1/\tau_2^C \quad [2]$$

$$\tau^S = \tau_1^S = 1/\tau_2^S \quad [3]$$

Questi vincoli garantiscono la condizione $\Pi \tau = 1$ vale a dire il prodotto dei τ^C per entrambi i livelli della variabile continuazione degli studi e dei τ^S per la variabile sesso sia uguale a 1.00.

$$\tau^{CS} = \tau_{11}^{CS} = \tau_{22}^{CS} = 1/\tau_{12}^{CS} = 1/\tau_{21}^{CS} \quad [4]$$

Dal momento che vi sono più parametri-effetto (9) che frequenze di cella (4), il modello saturo non potrebbe produrre stime senza i cinque vincoli sopra descritti dalle equazioni (2) (3) (4) sopra ricordate. Tali vincoli significano che solo quattro parametri-effetto sono indipendenti (η , C, S, CS).

Con quattro parametri-effetto indipendenti e quattro celle nella tabella, il modello saturo riprodurrà perfettamente le frequenze di cella osservate con 0 gradi di libertà residui (in generale il numero dei τ posti uguale a 1 determina i gradi di libertà).

Usando le equazioni della tabella CxS in Fig. 4 possiamo derivare le formule dei parametri τ in termini delle frequenze attese di cella. L'associazione tra continuazione e sesso è espressa dall'odds ratio atteso già calcolato in precedenza:

$$\Omega^{CS} = \text{odds ratio atteso (CS)} = \frac{F_{11} F_{22}}{F_{21} F_{12}} = \frac{F_{11}/F_{21}}{F_{12}/F_{22}} = 1.263$$

⁵ Ricordiamo che τ_1^C si riferisce al primo livello della variabile continuazione degli studi (continuano), τ_2^C al secondo livello (non continuano) e così per la variabile sesso.

Sostituendo alle varie frequenze attese le quattro equazioni della tabella CxS in Tabella 9 abbiamo:

$$\Omega^{CS} = \frac{F_{11} F_{22}}{F_{21} F_{12}} = \frac{(\eta \tau_1^C \tau_1^S \tau_{11}^{CS}) (\eta \tau_2^C \tau_2^S \tau_{22}^{CS})}{(\eta \tau_2^C \tau_2^S \tau_{21}^{CS}) (\eta \tau_1^C \tau_1^S \tau_{12}^{CS})} = \frac{\tau_{11}^{CS} \tau_{22}^{CS}}{\tau_{21}^{CS} \tau_{12}^{CS}}$$

Questa relazione mostra che l'odds ratio dipende esclusivamente dall'associazione fra C e S e non dalle distribuzioni marginali delle variabili. Usando le identità dell'equazione (4) possiamo riscrivere questi odds ratio in termini di una funzione di un singolo parametro a due variabili.

$$F_{11} F_{22} / F_{21} F_{12} = [\tau^{CS}]^4$$

$$\text{odds ratio} = [\tau^{CS}]^4$$

$$\text{oppure: } \tau^{CS} = (\text{odds ratio})^{1/4} = (F_{11} F_{22} / F_{21} F_{12})^{1/4}$$

che è la relazione già vista che collega l'odds ratio con il parametro τ^{CS} . Il parametro per la covarianza (continuazione x sesso) è la $4^{\sqrt{}}$ dell'odds ratio delle frequenze attese nel modello. Questo valore è 1.060.

Vediamo adesso la stima dei parametri ad una via τ_i^C e τ_j^S . Come suggeriscono Knoke e Burke, è bene partire dal prodotto dei due odds condizionati (Knoke-Burke, 1980, p. 15).

Infatti i parametri τ^C e τ^S rappresentano una misura dell'effetto ad una via delle variabili continuazione e sesso o, detto diversamente, una misura dell'asimmetria nella distribuzione di tali variabili. I parametri effetto τ rappresentano pertanto gli odds marginali medi di cella. Per poter attribuire alle singole celle della tabella questi parametri-effetto è necessario stimare gli odds marginali e poi considerarne la media (geometrica, dato il modello moltiplicativo) di cella. La miglior stima di questo effetto marginale è data

proprio dalla radice quarta degli 'odds products' che mettono a confronto le distribuzioni condizionate delle variabili considerate.

Nel nostro esempio, l'odds marginale continuano/non continuano è dato dal rapporto continuano (maschi + femmine) / non continuano (maschi + femmine).

La radice quarta di questo rapporto consente l'attribuzione della media geometrica alle singole celle della tabella.

Il parametro τ^c rappresenta la media geometrica dei due 'odds condizionati' sulla continuazione degli studi (cfr. D. Knoke, P.J. Burke, cit., p. 17).

$$\begin{aligned} \left(\frac{F_{11}}{F_{21}} \right) \left(\frac{F_{12}}{F_{22}} \right) &= \left(\frac{\eta \tau_1^c \tau_1^s \tau_{11}^{CS}}{\eta \tau_2^c \tau_1^s \tau_{21}^{CS}} \right) \left(\frac{\eta \tau_1^c \tau_2^s \tau_{12}^{CS}}{\eta \tau_2^c \tau_2^s \tau_{22}^{CS}} \right) \\ &= \left(\frac{\eta \tau_1^c \tau_{11}^{CS}}{\tau_2^c \tau_{21}^{CS}} \right) \left(\frac{\tau_1^c \tau_{12}^{CS}}{\tau_2^c \tau_{22}^{CS}} \right) = \frac{(\tau_1^c)^2}{(\tau_2^c)^2} = (\tau^c)^4 \end{aligned}$$

da cui

$$\tau^c = \left(\frac{F_{11} F_{12}}{F_{21} F_{22}} \right)^{1/4}$$

e, in modo analogo

$$\tau^s = \left(\frac{F_{11} F_{21}}{F_{12} F_{22}} \right)^{1/4}$$

Per i dati della nostra tabella si ha (come riportato analiticamente nella Tabella 10):

$$\tau^c = \left(\frac{\text{continuano}}{\text{non continuano}} (M) \cdot \frac{\text{continuano}}{\text{non continuano}} (F) \right)^{1/4} = (2.097 \cdot 1.660)^{1/4} = 1.366$$

$$\tau^s = \left(\frac{\text{maschi}}{\text{femmine}} (C_1) \cdot \frac{\text{maschi}}{\text{femmine}} (C_2) \right)^{1/4} = (1.086 \cdot 0.860)^{1/4} = 0.983$$

Un'altra espressione suggerita da Knoke e Burke indica che i coefficienti τ rappresentano il rapporto fra il numero di casi attesi in una categoria e la media geometrica dei casi attesi in tutte le categorie della Tabella 1.

Così

$$\tau_i^c = \frac{(F_{i1} F_{i2})^{1/2}}{(F_{11} F_{12} F_{21} F_{22})^{1/4}} = \frac{(F_{i1} F_{i2})^{1/2}}{\eta}$$

Con i nostri dati si ha:

$$\tau_1^c = \left(\frac{2185 \cdot 2012}{1535} \right)^{1/2} = 1.366$$

$$\tau_2^c = \left(\frac{1042 \cdot 1212}{1535} \right)^{1/2} = 0.732$$

e analogamente:

$$\tau_j^s = \frac{(F_{1j} F_{2j})^{1/2}}{\eta}$$

da cui:

$$\tau_1^S = \frac{(2185 \ 1042)^{1/2}}{1535} = 0.983$$

$$\tau_2^S = \frac{(2012 \ 1212)^{1/2}}{1535} = 1.017$$

Questa formula assicura il rispetto del vincolo $\Pi \tau = 1$. Tanto più il τ si discosta da 1.00 tanto più la distribuzione marginale si discosta dalla perfetta simmetria (distribuzione rettangolare nelle celle). Siamo in grado, a questo punto, di calcolare tutti i parametri stimati:

$$\eta = (f_{11} \ f_{12} \ f_{21} \ f_{22})^{1/4} = (2185 \ 2012 \ 1042 \ 1212)^{1/4} = 1.535$$

$$\tau_1^C = \frac{1}{\tau_2^C} = \frac{(f_{11} \ f_{12})^{1/2}}{\eta} = \frac{(2185 \ 2012)^{1/2}}{1535} = 1.366$$

$$\tau_1^S = \frac{1}{\tau_2^S} = \frac{(f_{11} \ f_{21})^{1/2}}{\eta} = \frac{(2185 \ 1042)^{1/2}}{1535} = 0.983$$

$$\tau_{11}^{CS} = \tau_{22}^{CS} = \frac{1}{\tau_{12}^{CS}} = \frac{1}{\tau_{21}^{CS}} = \left(\frac{f_{11} \ f_{22}}{f_{21} \ f_{12}} \right)^{1/4} = \left(\frac{2185 \ 1212}{1042 \ 2012} \right)^{1/4} = 1.060$$

Utilizzando queste stime provenienti dal modello saturo, possiamo riprodurre esattamente le quattro frequenze di cella:

	η	τ_1^C	τ_j^S	τ_{ij}^{CS}	F_{ij}
$F_{11} =$	1535	1.366	0.983	1.060	= 2185
$F_{12} =$	1535	1.366	1.017	0.943	= 2012
$F_{21} =$	1535	0.732	0.983	0.943	= 1042
$F_{22} =$	1535	0.732	1.017	1.060	= 1212

E' utile, a questo punto, richiamare il significato sostantivo di questi valori:

a) i soggetti che continuano gli studi dopo la terza media in provincia di Trento sono più numerosi di quelli che non continuano ($\tau^C = 1.360$);

b) la prevalenza dei maschi rispetto alle femmine nel gruppo di coloro che continuano è inferiore alla prevalenza delle femmine (rispetto ai maschi) nel gruppo di coloro che non continuano. Ciò spiega il valore del τ^S marginale inferiore all'unità ($\tau^S = 0.983$);

c) esiste una interazione fra sesso e continuazione degli studi nel senso che i maschi continuano più delle femmine ($\tau^{CS} = 1.060$). Tale relazione tuttavia non è simmetrica a causa delle asimmetrie nella variabile "continuazione degli studi" rispetto alla variabile "sesso".

Si ha infatti:

continuo	C_1	M 2185	} effetto medio di cella della variabile sesso
		F 2012 = 1.086	
non continuo	C_2	M 1042	} $\tau^S = (1.086 \ 0.860)^{1/4} = 0.983$
		F 1242 = 0.860	

In altri termini le femmine prevalgono sui maschi nel gruppo che non continua più di quanto i maschi prevalgono sulle femmine nel gruppo che continua.

L'apparente contraddizione emersa nei dati trova spiegazione nella logica sottostante la formula degli effetti τ ad una via. I parametri τ sono appunto la radice quadrata della media geometrica dei due odds condizionati nella continuazione degli studi. Essi sono diversi dagli odds marginali non condizionati calcolati per i marginali di riga della tabella e pari rispettivamente a: $4197/2254 = 1.862$ per gli odds marginali (C) e $3227/3224 = 1.001$ per gli odds marginali (S). Gli odds condizionati tengono conto della distribuzione dei casi delle altre variabili della tabella, mentre gli odds marginali non riflettono la presenza di altri fattori nei dati.

Il valore $\tau^s = 0.983$ non riflette tanto la prevalenza marginale delle femmine sui maschi (i maschi sono poco più numerosi delle femmine) quanto la maggior presenza delle femmine nel gruppo di coloro che non continuano rispetto alla maggiore presenza dei maschi nel gruppo di coloro che continuano. Questo effetto viene in evidenza solo nei casi in cui, come nel nostro esempio, la distribuzione marginale si avvicina alla distribuzione ottimale (50-50).

Per chiarire meglio questo punto consideriamo una tabella di dati artificiali in cui si abbia sempre la distribuzione maschi/femmine 50-50, sempre una prevalenza del gruppo di coloro che continuano rispetto a coloro che non continuano, ma che sia rovesciata la relazione interna degli odds condizionati nel senso che le femmine che continuano sono maggiori dei maschi che continuano (odds ratio condizionati = $60/90 = 0.667$) mentre, viceversa, nel gruppo di coloro che non continuano, il rapporto maschi/femmine è $40/10 = 4.0$.

La relazione fra sesso e continuazione degli studi esiste ed è fortemente asimmetrica data la forte asimmetria di una sola delle due variabili.

		[S]		
		S1	S2	Totale
[C]	C1	60	90	150
	C2	40	10	50
	Totale	100	100	200

In questa tabella i parametri sono i seguenti:

$$\eta = 38,3 \quad \tau_1^C = 1,917 \quad \tau_1^S = 1,278 \quad \tau_{11}^{CS} = 0,639$$

Concludendo, per avere dei τ ad una via tendenti al valore di 1.00 occorre che entrambe le variabili abbiano una distribuzione marginale perfettamente simmetrica. Tanto più ci si discosta da valori degli odds

marginali di 1.00, tanto più i parametri τ ad una via risentono di tale asimmetria.

Tabella 11 - I modelli moltiplicativi

Modello saturo:

$$\text{Frequenze } F_{ij} = \eta \tau_i^C \tau_j^S \tau_{ij}^{CS} \quad \eta = \text{antilog } \Theta$$

		F_{ij}	
		MASCHI	FEMMINE
Continuano	$F_{11} = \eta \tau_1^C \tau_1^S \tau_{11}^{CS}$	$F_{12} = \eta \tau_1^C \tau_2^S \tau_{12}^{CS}$	
Non continuano	$F_{21} = \eta \tau_2^C \tau_1^S \tau_{21}^{CS}$	$F_{22} = \eta \tau_2^C \tau_2^S \tau_{22}^{CS}$	

$$\eta = \sqrt[4]{F_{11} \cdot F_{12} \cdot F_{21} \cdot F_{22}} \quad (\text{media geometrica delle frequenze})$$

$$\tau_1^C = \sqrt[4]{\text{odds products}} = \left(\frac{F_{11}}{F_{21}} \cdot \frac{F_{12}}{F_{22}} \right)^{1/4} \quad \text{effetto C medio di cella}$$

$$\tau_j^S = \sqrt[4]{\text{odds products}} = \left(\frac{F_{11}}{F_{21}} \cdot \frac{F_{12}}{F_{22}} \right)^{1/4} \quad \text{effetto S medio di cella}$$

$$\tau_{ij}^{CS} = \sqrt[4]{\text{odds ratio}} = \left(\frac{F_{11}}{F_{21}} \cdot \frac{F_{12}}{F_{22}} \right)^{1/4} \quad \text{effetto CS medio di cella}$$

1.2.2. - Forma additiva del modello ⁶

Come abbiamo già ricordato, i modelli log-lineari additivi si presentano sotto due forme, sostanzialmente equivalenti⁷, espresse in termini di log-probabilità e log-frequenze.

I modelli additivi che operano rispettivamente sulle log-probabilità e sulle log-frequenze presentano la seguente forma generale (modello saturo):

LOG-PROBABILITA'

$$\log p_{ij} = V_{ij} = \mu + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS}$$

LOG-FREQUENZE

$$\log F_{ij} = V_{ij} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS}$$

Entrambi i modelli consistono di un termine medio, corrispondente all'idea di equiprobabilità, e di tre componenti additive per le quali valgono le seguenti relazioni:

$$\sum_i \lambda_i^C = \sum_j \lambda_j^S = \sum_i \lambda_{ij}^{CS} = \sum_j \lambda_{ij}^{CS} = 0$$

⁶ Per questa parte ci siamo rifatti, sostanzialmente, alla trattazione e alle formule e simboli di G.J.G. Upton (1978).

⁷ Si ha infatti:

$$y_{ij} = \log_e \left(\frac{f_{ij}}{f_{..}} \right) + \log_e (f_{..}) = \log p_{ij} + \text{costante}$$

Tali relazioni danno origine ai vincoli posti ai lambda del modello additivo saturo:

$$\lambda_2^C = -\lambda_1^C$$

$$\lambda_2^S = -\lambda_1^S$$

$$\lambda_{22}^{CS} = -\lambda_{12}^{CS} = -\lambda_{21}^{CS} = \lambda_{11}^{CS}$$

Nell'analisi dei modelli log-lineari per tabelle 2x2 l'analogia con l'analisi della varianza a due criteri è particolarmente evidente⁸ dato che ogni valore della tabella 2x2 può essere visto come espressione additiva degli elementi classici che compongono il teorema dell'analisi della varianza.

La formula generale della Anova esprime ogni punteggio individuale come somma di quattro componenti additive: media generale, eccedenza di riga, eccedenza di colonna e interazione.

$$x_{ij} = x_{..} + (x_{i.} - x_{..}) + (x_{.j} - x_{..}) + (x_{ij} - x_{i.} - x_{.j} + x_{..})$$

Per analogia si ha la seguente espressione delle singole probabilità di cella:

$$V_{ij} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS}$$

e applicando i simboli della tabella delle log-probabilità (cfr. Tab. n. 12):

$$V_{ij} = V_{..} + (V_{i.} - V_{..}) + (V_{.j} - V_{..}) + (V_{ij} - V_{i.} - V_{.j} + V_{..})$$

⁸ Nelder (1974), così come ricorda Upton (1978, p. 51), ha chiaramente mostrato come la maggior parte delle tecniche standard di analisi della varianza siano facilmente adattabili alla situazione dei log-lineari.

Anche nel caso del modello log-lineare ogni valore della tabella viene visto come composto da un effetto generale (Θ), da un effetto riga (λ^C), un effetto colonna (λ^S) e un effetto interazione (λ^{CS}).

I valori degli effetti additivi del modello sono pertanto i seguenti:

$$\lambda_i^C = V_{i.} - V_{..}$$

$$\lambda_j^S = V_{.j} - V_{..}$$

$$\lambda_{ij}^{CS} = V_{ij} - V_{i.} - V_{.j} + V_{..}$$

Sviluppando questi valori otteniamo, per $i = 1$:

$$\lambda_i^C = \frac{V_{11} + V_{12}}{2} - \frac{V_{11} + V_{12} + V_{21} + V_{22}}{4} = \frac{(V_{11} + V_{12}) - (V_{12} + V_{22})}{4} = \frac{(V_{11} - V_{21}) + (V_{12} - V_{22})}{4}$$

Questa espressione, al numeratore, non è altro che l'odds-products espresso in logaritmi come somma di differenze di logaritmi, perfettamente corrispondente al prodotto dei rapporti visto per i modelli moltiplicativi.

Anche la divisione per 4, diretta ad ottenere l'effetto medio di cella, corrisponde alla radice quadrata dei modelli moltiplicativi.

Analoga struttura ha la formula dell'effetto λ_j^S :

$$\lambda_j^S = \frac{(V_{11} - V_{12}) + (V_{21} - V_{22})}{4} = \frac{1}{4} (\text{odds products (S)})$$

Infine, l'effetto interazione λ_{ij}^{CS} ha la seguente formula:

$$\lambda_{ij}^{CS} = \frac{(V_{11} - V_{21}) - (V_{12} - V_{22})}{4} = \frac{1}{4} (\text{odds ratio (CS)})$$

Anche in questo caso il parallelismo in termini logaritmici con l'odds ratio medio di cella, già visto a proposito dei modelli moltiplicativi, è evidente. Si veda la Tabella n. 12.

Tabella n. 12 - Corrispondenza fra modelli additivi e moltiplicativi secondo la logica ANOVA (tab. 2x2).

MODELLO MOLTIPLICATIVO

EFFETTI A UNA VIA	EFFETTI INTERAZIONE	COSTANTE
$\tau_i^C ; \tau_j^S$	τ_{ij}^{CS}	η
odds	odds	
odds product	odds ratio	
(odds product) ^{1/4}	(odds ratio) ^{1/4}	

MODELLO ADDITIVO

$\lambda_i^C ; \lambda_j^S$	λ_{ij}^{CS}	Θ
odds product	odds ratio	
$\log (\text{odds product})^{1/4}$	$\log (\text{odds ratio})$	
$\frac{\log (\text{odds product})}{4}$	$\frac{\log (\text{odds ratio})}{4}$	

N.B.: $\lambda = \log \tau$

$\Theta = \log \eta$

1.2.2.1. - Modelli additivi sulle log-probabilità

Tornando ora alla Tabella 6 che riporta i dati del nostro esempio sulla continuazione degli studi per sesso, svolgiamo completamente i calcoli utilizzando il modello additivo e operando con le log-probabilità.

La Tabella n. 13, che riporta in modo analitico tutti i calcoli necessari per la stima dei parametri λ e τ , risulta di particolare utilità.

Dato il modello saturo:

$$V_{ij} = \mu + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS}$$

si ottengono i seguenti valori per le stime delle frequenze attese del modello saturo:

$$\mu = (-1.082 - 1.165 - 1.826 - 1.671)/4 = 1.436$$

$$\lambda_1^C = -1.123 + 1.436 = +0.313$$

$$\lambda_2^C = -1.748 + 1.436 = -0.313$$

$$\lambda_1^S = -1.454 + 1.436 = -0.018$$

$$\lambda_2^S = -1.418 + 1.436 = +0.018$$

$$\lambda_{11}^{CS} = -1.082 + 1.123 + 1.454 - 1.436 = +0.059$$

$$\lambda_{12}^{CS} = -1.165 + 1.123 + 1.418 - 1.436 = -0.059$$

$$\lambda_{21}^{CS} = -1.826 + 1.748 + 1.454 - 1.436 = -0.059$$

$$\lambda_{22}^{CS} = -1.671 + 1.748 + 1.418 - 1.436 = +0.059$$

Questi valori interpolano perfettamente i dati della tabella così come appare dallo schema seguente:

	μ	λ_i^C	λ_j^S	λ_{ij}^{CS}	V_{ij}	P_{ij}
$V_{11} =$	- 1.436	+ 0.313	- 0.018	+ 0.059	= - 1.082	0.339
$V_{12} =$	- 1.436	+ 0.313	+ 0.018	- 0.059	= - 1.165	0.312
$V_{21} =$	- 1.436	- 0.313	- 0.018	- 0.059	= - 1.826	0.161
$V_{22} =$	- 1.436	- 0.313	+ 0.018	+ 0.059	= - 1.671	0.188

L'interpretazione dei parametri appena calcolati è analoga a quella già vista per il modello moltiplicativo.

1.2.2.2. - Modelli additivi sulle log-frequenze

Finora abbiamo operato con i logaritmi delle probabilità di continuazione degli studi per sesso. Le cose non cambiano molto operando con i logaritmi delle frequenze:

$$y_{ij} = \log F_{ij}$$

In effetti la relazione fra log-frequenze e log-probabilità è di tipo lineare nel senso che:

$$y_{ij} = (\log p_{ij} - \text{costante})$$

Ne consegue che i due modelli sulle log-probabilità e sulle log-frequenze hanno identica struttura e differiscono soltanto per la costante che rappresenta l'effetto medio generale.

La forma del modello saturo con le log-frequenze è la seguente:

$$y_{ij} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS}$$

I valori λ stimati delle frequenze sono:

$$\lambda_i^C = y_{i.} - y_{..} = \frac{(y_{11} - y_{21}) + (y_{12} - y_{22})}{4}$$

$$\lambda_j^S = y_{.j} - y_{..} = \frac{(y_{11} - y_{12}) + (y_{21} - y_{22})}{4}$$

$$\lambda_{ij}^{CS} = y_{ij} - y_{i.} - y_{.j} + y_{..} = \frac{(y_{11} - y_{21}) - (y_{12} - y_{22})}{4}$$

$$\Theta = y_{..} = \log(F_{..})$$

Lo svolgimento dei calcoli, relativi al nostro esempio di Tabella 1, fornisce i seguenti valori:

$$\lambda_1^C = 7.648 - 7.336 = + 0.312$$

$$\lambda_1^S = 7.319 - 7.336 = - 0.017$$

$$\lambda_{11}^{CS} = 7.689 - 7.648 - 7.319 + 7.336 = 0.058$$

$$\Theta = 7.336$$

$$(\mu = - 1.386)$$

La stima delle frequenze attese prodotte dal modello saturo riproduce fedelmente, anche in questo caso, i valori osservati, così come risulta dal seguente prospetto:

	Θ	λ_i^C	λ_j^S	λ_{ij}^{CS}	y_{ij}	f_{ij}
y_{11}	7.336	+ 0.312	- 0.017	+ 0.058	7.689	2185
y_{12}	7.336	+ 0.312	+ 0.017	- 0.058	7.607	2012
y_{21}	7.336	- 0.312	- 0.017	- 0.058	6.949	1042
y_{22}	7.336	- 0.312	+ 0.017	+ 0.058	7.100	1212

I calcoli, date le ridotte dimensioni della tabella, sono stati condotti manualmente. Riportiamo comunque, relativamente al modello saturo, l'output del programma ECTA di L. Goodman che utilizzeremo regolarmente per i calcoli dei modelli più complessi. Come si noterà, il programma fornisce, oltre alle tavole delle frequenze osservate e attese, gli effetti λ del modello additivo e i corrispondenti parametri τ del modello moltiplicativo.

OUTPUT DEL PROGRAMMA ECTA PER TABELLE 2x2: MODELLO SATURO

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
4

NOTE
DISTRETTI PROSEGUIMENTO PER SESSO 2x2
VAR1=PROSEGUIMENTO/1=CONTINUA 2=NON CONTINUA
VAR2=SESSO/1=MASCHIO 2=FEMMINA

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
1 2 2

THE DIMENSION HAVE BEEN SET

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
2

INPUT DATA READ
2185.00 1042.0 2012.00 1212.00

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
4

NOTE
(CS) MODELLO SATURO

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
3 0 0 2 0 0 1 1 2 0 1 2

MARGINS FIT UNDER THE MODEL
1 2

AFTER ITERATION 2 THE LARGEST DEVIATION IS .00000
PEARSON .00, AND LIKELIHOOD RATIO 0.00 CHI-SQUARES
0 DEGREES OF FREEDOM

THE ORIGINAL TABLE, WITH THE FITTED VALUES SHOWN BENEATH IN EACH CELL
LEVELS OF VAR 2

VAR 1	1	2		
1	2185.00	2012.00	I	4197.00
	2185.00	2012.00	I	4197.00
2	1042.00	1212.00	I	2254.00
	1042.00	1212.00		2254.00
	3227.00	3224.00	I	6451.00
	3227.00	3224.00	I	6451.00

MARGINAL TABLE

	LEVELS OF VAR 1			
VAR 2	1	2		
1	2185.00	1042.00	I	3227.00
2	2012.00	1212.00	I	3224.00
	4197.00	2254.00	I	6451.00

THE ESTIMATED LAMBDA EFFECTS, THEIR STANDARD ERRORS, AND THE STANDARDIZED VALUES THESE THREE ARE FOLLOWED BY THE CORRESPONDING TAU PARAMETER OF THE MULTIPLICATIVE MODEL

VARIABLES OF ONLY TWO LEVELS WHERE THE SINGLE EFFECT SHOWN IS THE DIFFERENCE OF THE FIRST LEVEL, AND THE AVERAGE EFFECT

1 2	EFFECT	STD ERR	STOZD VAL	TAU PARM
	7.336			1535.014
EFFECT FOR VARIABLES				
1	.312	.013	23.831	1.366
2	-.017	.013	-1.311	.983
1 2	.058	.013	4.463	1.060

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
0

RUN COMPLETE 91 UNITS OF THE X DATA STORAGE ARRAY USED, OUT OF 4000

Tabella n. 13 - Modelli additivi: notazione per l'applicazione della formula (Tabelle 2x2)

Dati osservati:

F _{ij}			P _{ij}			
	M	F	Tot.	M	F	Tot.
C ₁	2185	2012	4197	.339	.312	.651
C ₂	1042	1212	2254	.161	.188	.349
Tot.	3227	3224	6451	.500	.500	1.000
				odds 2.105	1.660	

Simboli:

$$y_{ij} = \log F_{ij}$$

$$V_{.j} = \log P_{.j}$$

	S ₁	S ₂	Tot.	S ₁	S ₂	Tot.	$V_{i.} = \sum_j \frac{V_{ij}}{2}$
C ₁	y ₁₁	y ₁₂	y _{1.}	V ₁₁	V ₁₂	V _{1.}	
C ₂	y ₂₁	y ₂₂	y _{2.}	V ₂₁	V ₂₂	V _{2.}	
Tot.	y _{.1}	y _{.2}	y _{..}	V _{.1}	V _{.2}	V _{..}	

$$V_{.j} = \sum_i \frac{V_{ij}}{2}$$

Dati stimati:

	y _{ij}			V _{ij}			odds(S)	odds prod.	$\lambda^S = \frac{o/p}{4}$
	M	F	Tot.	M	F	Tot.			
C ₁	7.689	7.607	7.648	-1.082	-1.165	-1.123	+0.83		
C ₂	6.949	7.100	7.024	-1.826	-1.671	-1.748	-1.55	-0.72	-0.18
Tot.	7.319	7.353	7.336 (Θ)	-1.454	-1.418	-1.436 (μ) ¹			
odds (C) =	0.774	0.506		0.744	0.506				
odds pr. =	0.744+	0.506 =	1.250		1.250				
$\lambda^C = o.p/4 =$	0.313				0.313				
odds ratio =	0.238				0.238				
$o.r./4 - \lambda^{CS} =$	0.059				0.059				

¹ N.B.: il logaritmo di una somma non è la somma dei logaritmi ma la loro media geometrica

(segue)

Tabella 13: (segue)

$$\begin{aligned} \lambda^C &= y_{i.} - y_{..} = 7.648 - 7.336 = 0.313 & \lambda^C &= V_{i.} - V_{..} = -1.123 + 1.436 = 0.313 \\ \lambda^S &= y_{.j} - y_{..} = 7.319 - 7.336 = -0.018 & \lambda^S &= V_{.j} - V_{..} = -1.454 + 1.436 = -0.018 \\ \lambda^{CS} &= y_{11} - y_{1.} - y_{.1} + y_{..} = & \lambda^{CS} &= V_{11} - V_{1.} - V_{.1} + V_{..} = \\ &= 7.689 - 7.648 - 7.319 + 7.336 = 0.059 & &= 1.082 + 1.123 + 1.454 - 1.436 = 0.059 \end{aligned}$$

	τ_{ij}			$\lambda_{ij} = \log \tau_{ij}$		
	M	F	Tot.	M	F	Tot.
C ₁	1.060	.943	1.366	+.059	-.059	+.313
C ₂	.943	1.060	.732	-.059	+.059	-.313
Tot.	.938	1.017	1.535	-.018	+.018	-1.436

Calcolo dei parametri per i modelli log-lineari 2x2x2

Modello moltiplicativo

$$\eta = 622.65$$

$$\tau_1^C = \left(\frac{F_{111}}{F_{211}} \cdot \frac{F_{121}}{F_{221}} \right) \left(\frac{F_{112}}{F_{212}} \cdot \frac{F_{122}}{F_{222}} \right)^{1/8} \quad (\text{prodotto dei due odds products parziali CS/T})$$

$$\tau_1^C = \left(\frac{1246}{726} \cdot \frac{1111}{874} \cdot \frac{824}{185} \cdot \frac{823}{205} \right)^{1/8} = 1.581$$

oppure (formula abbreviata):

$$\tau_1^C = (F_{111} F_{121} F_{112} F_{122})^{1/4} / \eta = (1246 \cdot 1111 \cdot 824 \cdot 823)^{1/4} / 622.65 = 1.581$$

$$\tau_1^S = \left(\frac{F_{111}}{F_{121}} \cdot \frac{F_{211}}{F_{221}} \cdot \frac{F_{112}}{F_{122}} \cdot \frac{F_{212}}{F_{222}} \right)^{1/8} = (\text{prodotto dei due odds product parziali SC/T})$$

$$= \left(\frac{1246}{1111} \cdot \frac{726}{874} \cdot \frac{824}{823} \cdot \frac{185}{205} \right)^{1/8} = 0.979$$

(segue)

Tabella 13: (segue)

oppure (formula abbreviata)

$$\tau_1^S = (F_{111} \cdot F_{211} \cdot F_{112} \cdot F_{212})^{1/4} / \eta = (1246 \cdot 726 \cdot 185)^{1/4} / 622.65 = 0.979$$

$$\tau_1^T = \left(\frac{F_{111}}{F_{112}} \cdot \frac{F_{211}}{F_{212}} \cdot \frac{F_{121}}{F_{122}} \cdot \frac{F_{221}}{F_{222}} \right)^{1/8} \quad (\text{Prodotto dei due odds products parziali CT/S})$$

$$= \left(\frac{1246}{824} \cdot \frac{726}{185} \cdot \frac{1111}{823} \cdot \frac{874}{205} \right)^{1/8} = 1.555$$

oppure (formula abbreviata):

$$\tau_1^T = (1246 \cdot 1111 \cdot 726 \cdot 874)^{1/4} / \eta = 1.555$$

$$\tau_{11}^{CS} = \left(\frac{F_{11.}}{F_{21.}} \cdot \frac{F_{.12}}{F_{.22}} \right)^{1/8} = \left(\frac{2070}{911} \cdot \frac{1934}{1079} \right)^{1/8} = 1.030 \quad (\text{odds ratio marginale (T)})$$

$$\tau_{1.1}^{CT} = \left(\frac{F_{1.1}}{F_{2.1}} \cdot \frac{F_{.12}}{F_{.2.2}} \right)^{1/8} = \left(\frac{2357}{1600} \cdot \frac{1647}{390} \right)^{1/8} = 0.877$$

$$\tau_{.11}^{ST} = \left(\frac{F_{.11}}{F_{.21}} \cdot \frac{F_{.12}}{F_{.22}} \right)^{1/8} = \left(\frac{1972}{1985} \cdot \frac{1009}{1028} \right)^{1/8} = 1.002$$

$$\tau_{111}^{CST} = \left(\frac{F_{111}}{F_{211}} \cdot \frac{F_{121}}{F_{221}} / \frac{F_{112}}{F_{212}} \cdot \frac{F_{122}}{F_{222}} \right)^{1/8} = 1.025$$

N.B. $\tau_{111}^{CTS} = (\text{odds ratio})^{1/8}$
 $1.025 = (1.217)^{1/8}$

(segue)

Tabella 13 (segue)

Modello additivo:

$$\Theta = y_{...} = \frac{\Sigma \Sigma \Sigma y_{ijk}}{8} = 6.434$$

$$\lambda_i^C = y_{i..} - y_{...} = 6.892 - 6.434 = + 0.458$$

$$\lambda_j^S = y_{.j.} - y_{...} = 6.412 - 6.434 = - 0.022$$

$$\lambda_k^T = y_{..k} - y_{...} = 6.875 - 6.434 = + 0.041$$

$$\lambda_{ij}^{CS} = y_{ij.} - y_{i..} + y_{...} = 6.921 - 6.892 - 6.412 + 6.434 = 0.051$$

$$\lambda_{ik}^{CT} = y_{i.k} - y_{i..} - y_{..k} + y_{...} = 7.070 - 6.892 - 6.875 + 6.434 = - 0.263$$

$$\lambda_{jk}^{ST} = y_{.jk} - y_{.j.} - y_{..k} + y_{...} = 6.858 - 6.412 - 6.875 + 6.434 = + 0.005$$

$$\lambda_{ijk}^{CST} = y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...} = 7.128 - 6.921 - 7.070 - 6.858 + 6.892 + 6.412 + 6.875 - 6.434 = 0.024$$

Calcolo delle frequenze stimate (modello saturo)

Dal modello saturo moltiplicativo:

$$f_{ijk} = \eta \cdot \tau^C \cdot \tau^S \cdot \tau^T \cdot \tau^{CS} \cdot \tau^{CT} \cdot \tau^{ST} \cdot \tau^{CST}$$

$$f_{111} = 622.65 \cdot 1.581 \cdot 0.979 \cdot 1.555 \cdot 1.052 \cdot 0.769 \cdot 1.004 \cdot 1.025 = 1.246$$

.....

Dal modello saturo additivo:

$$y_{ijk} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_k^T + \lambda_{ij}^{CS} + \lambda_{ik}^{CT} + \lambda_{jk}^{ST} + \lambda_{ijk}^{CST}$$

$$y_{111} = 6.434 + 0.458 - 0.022 + 0.441 + 0.051 - 0.263 + 0.005 + 0.024 = 7.128 \rightarrow f_{111} = 1.246$$

.....

1.3. - Corrispondenza fra modelli additivi e moltiplicativi

Abbiamo già visto in varie parti di questa trattazione come i vari elementi dei modelli log-lineari additivi e moltiplicativi si corrispondono puntualmente essendo entrambi centrati sui concetti di odds, odds marginali, odds condizionati, odds ratio, odds products, differendo unicamente per la forma dell'espressione dei valori della tabella 2x2 di partenza.

In questo paragrafo cerchiamo di fornire una rassegna sistematica delle corrispondenze tra tutti gli elementi del modello.

Per quanto riguarda la forma generale del modello saturo si hanno le seguenti equazioni:

Modelli additivi

$$y_{ij} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS} \text{ per log-frequenze}$$

$$v_{ij} = \mu + \lambda_i^C + \lambda_j^S + \lambda_{ij}^{CS} \text{ per log-probabilità}$$

per quanto riguarda l'effetto generale si ha:

$$\Sigma \lambda_{ij} = 0$$

$$\Theta = y_{..} = \frac{\Sigma \log F_{ij}}{4}$$

$$\mu = V_{..} = \frac{\Sigma \log p_{ij}}{4}$$

Modelli moltiplicativi

$$F_{ij} = \eta \tau_i^C \tau_j^S \tau_{ij}^{CS} \text{ per frequenze}$$

$$p_{ij} = \mu \tau_i^C \tau_j^S \tau_{ij}^{CS} \text{ per probabilità}$$

$$\prod \tau_{ij} = 1$$

$$\eta = (\prod F_{ij})^{1/4}$$

$$\mu = (\prod p_{ij})^{1/4}$$

Gli effetti marginali a una via (identici sia nel caso delle probabilità che delle frequenze) sono i seguenti:

per log-frequenze

$$\lambda_i^C = y_{i.} - y_{..} = \frac{(y_{11}-y_{21})+(y_{12}-y_{22})}{4}$$

$$\lambda_j^S = y_{.j} - y_{..} = \frac{(y_{11}-y_{12})+(y_{21}-y_{22})}{4}$$

per log-probabilità

$$\lambda_i^C = V_{i.} - V_{..} = \frac{(V_{11}-V_{12})+(V_{21}-V_{22})}{4}$$

$$\lambda_j^S = V_{.j} - V_{..} = \frac{(V_{11}-V_{12})+(V_{21}-V_{22})}{4}$$

per frequenze

$$\tau_i^C = \left(\frac{F_{11}}{F_{21}} \frac{F_{12}}{F_{22}} \right)^{1/4}$$

$$\tau_j^S = \left(\frac{F_{11}}{F_{12}} \frac{F_{21}}{F_{22}} \right)^{1/4}$$

per probabilità

$$\tau_i^C = \left(\frac{P_{11}}{P_{21}} \frac{P_{12}}{P_{22}} \right)^{1/4}$$

$$\tau_j^S = \left(\frac{P_{11}}{P_{12}} \frac{P_{21}}{P_{22}} \right)^{1/4}$$

Per quanto riguarda, infine, gli effetti interazione, abbiamo le seguenti espressioni:

$$\lambda_{ij}^{CS} = y_{ij} - y_{i.} - y_{.j} + y_{..} = \frac{(y_{11}-y_{21}) - (y_{21}-y_{22})}{4} \quad \tau_{ij}^{CS} = \left(\frac{F_{11}}{F_{21}} : \frac{F_{12}}{F_{22}} \right)^{1/4}$$

$$\tau_{ij}^{CS} = V_{ij} - V_{i.} - V_{.j} + V_{..} = \frac{(V_{11}-V_{21}) - (V_{21}-V_{22})}{4} \quad \tau_{ij}^{CS} = \left(\frac{P_{11}}{P_{21}} : \frac{P_{12}}{P_{22}} \right)^{1/4}$$

I modelli log-lineari additivi e moltiplicativi, come abbiamo già osservato, si rifanno sia all'approccio degli odds (e log-odds) sia all'approccio dell'analisi della varianza.

E' utile quindi osservare le formule con questa duplice ottica:

Modelli additivi

Anova

$$\lambda_i^C \text{ (effetto riga)} = (\text{media marginale (i)} - \text{media generale})$$

$$\lambda_j^S \text{ (effetto colonna)} = (\text{media marginale (j)} - \text{media generale})$$

$$\lambda_{ij}^{CS} \text{ (effetto interazione)} = (\text{valore di cella} - \text{media di riga} - \text{media di colonna} + \text{media generale})$$

Odds

$$\lambda_i^C = \frac{\text{odds product (C)}}{4} = \frac{\text{odds product (C)}}{\text{medio di cella}}$$

$$\lambda_j^S = \frac{\text{odds product (S)}}{4} = \frac{\text{odds product (S)}}{\text{medio di cella}}$$

$$\lambda_{ij}^{CS} = \frac{\text{odds ratio (CS)}}{4} = \frac{\text{odds ratio}}{\text{medio di cella}}$$

Modelli moltiplicativi

Anova

$$\tau_i^C \text{ (effetto riga)} = F_{1.} / F_{..}$$

$$\tau_j^S \text{ (effetto colonna)} = F_{.j} / F_{..}$$

$$\tau_{ij}^{CS} \text{ (effetto interazione)} = F_{11} : F_{1.} / F_{.j} : F_{..}$$

Odds

$$\tau_i^C = (\text{odds product (C)})^{1/4}$$

$$\tau_j^S = (\text{odds product (S)})^{1/4}$$

$$\tau_{ij}^{CS} = (\text{odds ratio (CS)})^{1/4}$$

La Tabella n. 14 riporta i valori dei parametri calcolati sui dati della nostra tabella rispettivamente per il modello additivo e moltiplicativo.

Tabella 14 - Corrispondenza fra modelli log-lineari additivi e moltiplicativi

	log _e F _{ij}	F _{ij}
variabili (frequenze di celle attese)	y ₁₁ = 7.689	F ₁₁ = 2.185
effetto generale	Θ = 7.336	η = 1.535
effetti marginali	λ _i ^C = + 0.313	τ _i ^C = 1.366
	λ _j ^S = - 0.018	τ _j ^S = 0.983
effetti interazione	λ _{ij} ^{CS} = + 0.059	τ _{ij} ^{CS} = 1.060

Corrispondenza fra odds, τ, λ (1)

Marginal odds	Conditional odds	Odds ratio	Odds product	τ ^{CS} = √ ⁴ odds ratio	τ ^C	τ ^S	λ ^{CS} = log(odds ratio)/4	λ ^C	λ ^S
C 1.862	C/M = 2.097	1.263	3.481	√ ⁴ 1.263 = 1.060	√ ⁴ 3.481 = 1.366		log(1.263)/4 = 0.059	log(3.481)/4 = 0.312	
	C/F = 1.660								
S 1.001	S/C1 = 1.806	1.263	0.934			√ ⁴ 0.394 = 0.983			log(0.934)/4 = 0.01
	S/C2 = 0.860								

N.B.: I τ e i λ sono odds ratio medi di cella, il primo (τ), espresso in forma moltiplicativa, è media geometrica; il secondo (λ), espresso in forma additiva, è media aritmetica.

1.4. - Altri modelli per tabelle 2x2: i modelli non saturi

Il modello saturo, fin qui esaminato, contiene tanti parametri quante sono le celle nella tabella e di conseguenza adatta in modo perfetto le frequenze di cella. Tuttavia, obiettivo del ricercatore è quello di definire modelli di spiegazione dei dati relativamente economici, o "parsimoniosi". Passiamo quindi in rassegna alcuni modelli più semplici che rinunciano ad alcuni parametri presenti nel modello saturo.

La scelta del modello più efficiente viene operata attraverso il test chi-quadrato per bontà di adattamento calcolato sulle differenze fra frequenze osservate e frequenze attese dal modello.

La soppressione di un parametro, che equivale a porlo uguale a 1.00 nei modelli moltiplicativi, oppure uguale a 0.00 nei modelli additivi, rende molto simili le procedure del modello log-lineare a quelle della regressione multipla. Naturalmente i modelli non saturi, privi di qualche effetto-parametro, producono frequenze di cella attese diverse dai dati osservati. Un primo modello da prendere in considerazione è il *modello di indipendenza*. La sua forma generale, nel caso di log-probabilità in forma additiva, è la seguente:

$$V_{ij} = \mu + \lambda_i^C + \lambda_j^S$$

Il modello di indipendenza risulta pertanto privo dell'effetto interazione: λ^{CS}, fondandosi sull'ipotesi che le frequenze attese dipendono esclusivamente dai marginali di riga e di colonna oltre che dall'effetto generale (μ).

Altri modelli più semplici, che ipotizzano un solo effetto ad una via, sono i modelli che presuppongono:

Nessun effetto di C: $V_{ij} = \mu + \lambda_j^S$

Nessun effetto di S: $V_{ij} = \mu + \lambda_i^C$

Equiprobabilità: $V_{ij} = \mu$

Data una tabella 2x2 vi sono pertanto cinque modelli possibili legati fra loro da una relazione gerarchica secondo la quale un modello che contiene un parametro di ordine superiore include automaticamente tutti i parametri di ordine inferiore ed esclude quelli di ordine superiore.

Un modello non gerarchico, viceversa, è un modello che include solo un effetto interazione senza i corrispondenti effetti ad una via, ad esempio:

$$V_{ij} = \mu + \lambda_{ij}^{CS}$$

La Tabella n. 15 riporta la costruzione dei cinque modelli usando i dati della nostra ricerca; in essa sono riportati analiticamente i dati delle frequenze attese nelle varie ipotesi e le stime dei parametri oltre, naturalmente, ai valori di L^2 (likelihood ratio) e del suo standard error⁹.

Al fine di costruire il "miglior modello" che adatta le frequenze osservate della nostra tabella di contingenza e quindi per valutare la significatività dei singoli parametri-effetto, la procedura più conveniente, con tabelle di dimensioni ridotte, è quella di procedere a confronti a coppie fra modelli diversi.

Nel caso del nostro esempio gli effetti da definire sono C, S, CS. Per testare la significatività dell'effetto S procediamo al confronto fra il modello [2] di equiprobabilità (C)(S) che comprende gli effetti di μ , λ^C e λ^S e il modello [3] di nessun effetto di S (C), che comprende solo gli effetti μ e λ^C .

La differenza fra i due modelli, definita in termini di riduzione di L^2 ottenuta, con gradi di libertà pari alla differenza fra i gradi di libertà dei due

⁹ Le formule di L^2 e di G^2 , adattate da G.J.G. Upton, (1978, p. 25), sono le seguenti:

$$L^2 = 2 \sum \sum f_{ij} \log \left(\frac{f_{ij}}{e_{ij}} \right)$$

$$G^2 = 2 \left[\sum F_{ij} (\log e_{ij} - \log F_{ij}) \right]$$

modelli messi a confronto, mostra che il fattore S non è significativo (l'effetto della distribuzione fra i maschi e le femmine sulla configurazione della tavola di contingenza è nullo).

Per testare la significatività dell'effetto C la differenza prende in considerazione i modelli [2] di indipendenza (C)(S) e [4] di nessun effetto di C (S).

Come mostrano i dati della Tabella n. 16.A. l'inserimento nel modello del fattore C risulta fortemente significativo. Questo risultato era ormai scontato data la notevole asimmetria fra il numero di coloro che continuano e non continuano nella tabella.

Il fattore interazione viene testato attraverso la differenza fra il modello saturo (CS) che comprende tutti gli effetti λ^C , λ^S , λ^{CS} e il modello di indipendenza che manca, appunto, di quest'ultimo effetto interazione. La differenza fra i due modelli evidenzia un miglioramento nell' L^2 di 19.46, differenza che risulta significativa con due gradi di libertà allo 0.01.

I risultati di questi confronti permettono la definizione del modello finale che sarà:

$$F_{ij} = \eta \cdot \tau_i^C \cdot \tau_{ij}^{CS} \quad \text{oppure} \quad \log (F_{ij}) = \Theta + \lambda_i^C + \lambda_{ij}^{CS}$$

$$F_{ij} = 1535 \cdot 1.366 \cdot 1.060 \quad \log (F_{ij}) = 7.336 + 0.312 + 0.058$$

Questo modello esprime la struttura interna dei dati della nostra tabella: *il numero degli studenti che continuano gli studi dopo la terza media in provincia di Trento è notevolmente superiore di quello di coloro che non continuano. Non risultano apprezzabili differenze nel numero complessivo di maschi e di femmine presenti nel sistema scolastico alla fine del ciclo dell'obbligo (3^a media).*

Esiste un effetto interazione fra continuazione degli studi e sesso nel senso che la proporzione di coloro che continuano è superiore nei maschi rispetto alle femmine le quali, viceversa, abbandonano in misura superiore

ai maschi dopo la 3^a media.

Un modo alternativo per valutare il contributo dei singoli parametri alla significatività del modello, come fa osservare Upton (1978, pp. 56-57), è quello di procedere in direzione FORWARD, partendo dal modello più semplice di equiprobabilità ed aggiungendovi, ad uno ad uno, tutti i parametri effetto fino ad arrivare al modello saturo. Tale procedura è illustrata nella Tabella n. 16.B.

Tabella 15 - Modelli log-lineari alternativi per i dati della tabella 2x2. Proseguimento degli studi per sesso in provincia di Trento.

Frequenze attese secondo i vari modelli

	Equiprobabilità			Nessun effetto di S C			Nessun effetto di C S			Indipendenza C S			Saturo C S		
	F _{ij} = η			F _{ij} = η τ _i ^C			F _{ij} = η τ _j ^S			F _{ij} = η τ _i ^C τ _j ^S			F _{ij} = η τ _i ^C τ _j ^S τ _{ij} ^{CS}		
C1	1613	1613	3226	2098.5	2098.5	4197	1613.5	1612	3225.5	2099.5	2097.5	4197	2185	2012	4197
C2	1613	1613	3226	1127.0	1127.0	2254	1613.5	1612	3225.5	1127.5	1126.5	2254	1042	1212	2254
Tot.	3226	3226	6451	3225.5	3225.5	6451	3227.0	3224	6451.0	3227.0	3224.0	6451	3227	3224	6451
	χ ² = 603.36			χ ² = 19.95			χ ² = 603.36			χ ² = 19.94			χ ² = 0.0		
	L ² = 614.37			L ² = 19.97			L ² = 614.37			L ² = 19.96			L ² = 0.0		

Risultati dei modelli alternativi per i dati della tabella n. 1

Modello	e ₁₁	e ₁₂	e ₂₁	e ₂₂	λ ₁ ^C	λ ₁ ^S	λ ₁₁ ^{CS}	μ	g.l.	L ²	SE(λ)	λ ₁ ^C /SEλ	λ ₁ ^S /SEλ	λ ₁₁ ^{CS} /SEλ
1 Saturo (C S)	2185	2012	1042	1212	+0.312	-0.017	+0.058	7.336	0	0	0.013	24	1.3	4.46
2 Indipendenza (C) (S)	2099.5	2097.5	1127.5	1126.5	+0.286	+0.001	+0.000	7.338	1	19.96	0.228	1.650	0.01	0.0
3 Nessun effetto S (C)	2098.5	2098.5	1127	1127	+0.311	0	0	7.338	2	19.97	0.228	1.364	0	0
4 Nessun effetto C (S)	1613.5	1612	1613.5	1612	0	0.002	0	7.386	2	614.37	0.223	0	0.01	0
5 Equiprobabilità ()	1613	1613	1613	1613	0	0	0	7.386	3	614.37	0.223	0	0	0

N.B.:

τ = antilog λ

$$S.E. (\lambda) = \sqrt{\frac{\sum \sum \frac{1}{F_{ij}}}{(n.celle)^2}} = 0.013$$

Tabella 16 - Confronto fra effetti di diversi modelli in tabelle 2x2

16.A

Modello	Parametri inclusi	g.l.	L ²	S.E. (L ²)
2 (C) (S)	$\hat{\mu}, \lambda_1^C, \lambda_1^S$	1	19.96	
3 (C)	$\hat{\mu}, \lambda_1^C$	2	19.97	
Differenze	λ_1^S	1	0.01	Il fattore S non è significativo
2 (C) (S)	$\hat{\mu}, \lambda_1^C, \lambda_1^S$	1	19.26	
4 (C)	$\hat{\mu}, \lambda_1^S$	2	614.37	
Differenze	λ_1^C	1	594.41	** Il fattore C è significativo
1 (C S)	$\hat{\mu}, \lambda^C, \lambda^S, \lambda^{CS}$	0	0	
2 (C) (S)	$\hat{\mu}, \lambda^C, \lambda^S$	2	19.96	
Differenze	λ^{CS}	2	10.96	** Il fattore interazione CS è significativo

16.B

Modello	Parametri inclusi	g.l.	L ²	Riduzione di L ²
5 equiprobabilità	μ	3	614.37	
4 nessun effetto C (S)	μ, λ_1^S	2	614.37	0.0
3 nessun effetto S (C)	μ, λ_1^C	2	19.97	594.41 **
2 indipendenza (C) (S)	$\mu, \lambda_1^S, \lambda_1^C$	1	19.96	0.01
1 saturo (CS)	$\mu, \lambda_1^S, \lambda_1^C, \lambda_{11}^{CS}$	0	0	19.96

Nel capitolo precedente abbiamo visto in dettaglio tutta l'informazione deducibile dai dati di una tabella 2x2. Malgrado la semplicità di tale struttura, come abbiamo visto, per interpretare i dati possono essere definiti ben cinque modelli.

Quando le dimensioni della tabella crescono, il numero di modelli (gerarchici e non) che da essa si possono derivare aumenta in modo esponenziale. Goodman (1970) mostra come questo numero diventa 19 per tabelle a tre variabili dicotomiche, 167 per tabelle a quattro variabili dicotomiche e sale ad alcune migliaia per tabelle a cinque variabili.

Come orientarsi in questa selva di modelli alternativi? Senza alcuna guida il compito risulterebbe oltremodo difficile se non impossibile. Al di là della selezione dei modelli su basi teoriche, indispensabile per questa operazione, esiste una guida di tipo statistico costituita dal modello saturo spesso usato come valido punto di partenza per l'analisi (Upton, 1978, p. 59).

Quando adattiamo il modello saturo stimiamo i valori di tutti i λ (il linguaggio dei modelli additivi è estendibile anche per quelli moltiplicativi) che potranno poi essere inclusi in modelli più semplici.

I valori λ vicino allo 0 o, in alternativa, i valori τ prossimi a 1 sono, di regola, di scarsa importanza e pertanto la scelta dei modelli non saturi sarà guidata dall'obiettivo di includere quei λ che deviano apprezzabilmente dallo 0 o i valori τ che maggiormente si discostano dal valore 1.

2.1. - La tabella generale a tre dimensioni con variabili dicotomiche

Vediamo innanzitutto di esprimere in termini generali il modello saturo che interpreta in modo completo le quantità V_{ijk} (nel caso di log-probabilità) o y_{ijk} (nel caso di log-frequenze) in termini di media generale, effetti principali delle singole variabili C (continuazione degli studi), S (sesso), T (titolo di studio del padre); le tre interazioni a due variabili (o

interazioni di primo ordine) CS, CT, ST e l'interazione a tre variabili (o di secondo ordine) CST.

Introduciamo a questo punto alcune notazioni convenzionali usate per descrivere i modelli in modo sintetico, senza fare riferimento all'intera distribuzione delle frequenze di cella.

Abbiamo già accennato al concetto di struttura gerarchica, che per le tavole 2x2 è di facile e immediata comprensione.

Nelle tabelle multidimensionali, come si è già detto, i modelli di adattamento ai dati osservati crescono rapidamente di numero e di complessità. Di conseguenza vale la pena di chiarire ulteriormente questo aspetto.

Si ha una struttura gerarchica di modelli quando una complessa relazione multivariata presente nei dati deve essere scomposta in una serie di relazioni più semplici gerarchicamente ordinate. Ad esempio, nella classificazione a 3 variabili Continuazione/Sesso/Titolo di studio del padre, un modello gerarchico comprendente l'interazione di secondo ordine fra continuazione, sesso e titolo di studio include necessariamente tutti i parametri effetto-interazione di primo ordine (continuazione/sesso; continuazione/titolo; sesso/titolo). Esprimendo queste relazioni nel modo simbolico solitamente utilizzato avremo che nei modelli gerarchici l'effetto di secondo ordine τ^{CST} include (in modo, appunto, gerarchico) gli effetti di 1° ordine τ^{CS} , τ^{CT} , τ^{ST} , così come gli effetti ad una via τ^C , τ^S , τ^T e, naturalmente, l'effetto generale η . In altre parole, in una struttura gerarchica i modelli che contengono relazioni di ordine superiore includono implicitamente tutti i possibili effetti di ordine inferiore (Knoke, Burke, 1980, p. 20). I modelli log-lineari comprendono sia gli approcci gerarchici sia quelli non gerarchici anche se i primi sono generalmente preferiti.

Per quanto riguarda la notazione sintetica dei vari modelli si utilizzano solitamente lettere che indicano le variabili fra loro in relazione nel modello racchiuse fra parentesi. Ciascun insieme di lettere entro parentesi identifica il parametro effetto dell'ordine più elevato incluso nel modello e, automaticamente, in virtù della struttura gerarchica, tutte le relazioni di

ordine inferiore necessariamente presenti.

Ad esempio, il modello saturo per i dati della Tabella 17, già scritto con l'espressione

$$F_{ijk} = \eta \cdot \tau^C \cdot \tau^S \cdot \tau^T \cdot \tau^{CS} \cdot \tau^{CT} \cdot \tau^{ST} \cdot \tau^{CST}$$

si indica con (CST). Tale notazione implica che nel modello sono automaticamente compresi i modelli a due vie (CS), (CT), (ST) i quali, a loro volta, includono in via gerarchica i parametri singoli C, S, T in essi indicati.

La puntuale corrispondenza fra le notazioni dei vari modelli a 3 variabili è contenuta nel seguente prospetto riportato da Upton e da lui indicato con il termine di "TALLY CHART" (Upton, 1978, p. 72).

Mapa dei possibili modelli gerarchici a tre variabili

	Modelli	Parametri							
		η	τ^C	τ^S	τ^T	τ^{CS}	τ^{CT}	τ^{ST}	τ^{CST}
(Saturo)	1. CST	x	x	x	x	x	x	x	x
(Senza interazione)	2. CS CT ST	x	x	x	x	x	x	x	x
(Indipendenza)	3. CS CT	x	x	x	x	x	x		
	4. CS ST	x	x	x	x	x		x	
	5. CT ST	x	x	x	x		x	x	
	6. C ST	x	x	x	x			x	
	7. S CT	x	x	x	x		x		
	8. T CS	x	x	x	x	x			
	9. ST	x		x	x			x	
	10. CT	x	x		x		x		
	11. CS	x	x	x		x			
	12. C S T	x	x	x	x				
	13. C S	x	x	x					
	14. C T	x	x		x				
	15. S T	x		x	x				
	16. C	x	x						
	17. S	x		x					
	18. T	x			x				
(Equiprobabilità)	19. η	x							

Oltre alla compattezza e alla sintesi, la notazione standard esprime un altro importante aspetto dell'analisi. Le variabili incluse nelle parentesi identificano le sottotavole (o tabelle marginali) che derivano dalla completa tabulazione incrociata.

Nella stima delle frequenze attese di una tabella generale, secondo le ipotesi incluse in un dato modello, le frequenze attese nelle singole tabelle marginali devono eguagliare le corrispondenti frequenze osservate. In questo senso la notazione standard è spesso chiamata notazione di "adattamento dei marginali".

Come è già illustrato nella Tabella 15 che riporta i vari modelli possibili per tabelle 2x2, anche se le singole frequenze di cella previste nelle varie ipotesi si discostano da quelle osservate, l'uguaglianza fra frequenze attese e osservate è sempre presente nelle distribuzioni marginali.

Rimane il problema pratico di ottenere le frequenze attese secondo i diversi modelli ipotizzati. Per modelli semplici del tipo (2x2), come abbiamo mostrato nel paragrafo precedente, o del tipo (2x2x2), come avremo modo di vedere più avanti, esistono formule relativamente semplici che permettono di stimare le frequenze attese per i vari modelli saturi e non saturi.

Per modelli più complessi è necessario ricorrere ad algoritmi che generano le frequenze attese. Oltre all'algoritmo di Stephan-Deming utilizzato dal programma ECTA di Fay e Goodman (cfr. P. Corbetta, H. Schadee, 1984), esiste l'algoritmo di Newton Raphson utilizzato nel programma MULTIQUAL di Bock-Yates.

La procedura iterativa di Stephan-Deming, come è stato accennato in precedenza nel Capitolo 1.1., genera le stime di massima verosimiglianza delle frequenze attese di cella per un modello gerarchico. Le stime delle frequenze attese ottenute in via preliminare vengono successivamente aggiustate per adattarsi a ciascuna sottotavola marginale specificata nel modello.

I valori iniziali assegnati alle frequenze sono pari a 1 e la conversione entro le stime finali è estremamente rapida. Ad ogni passo del processo iterativo le vecchie stime vengono sostituite con le nuove in una

successione di cicli che termina al raggiungimento di una soglia assegnata di miglioramento delle stime stesse.

Il programma ECTA, utilizzando questo algoritmo, produce le frequenze attese per ciascun modello ipotizzato e automaticamente fornisce le stime dei parametri effetto (τ e λ nelle due versioni moltiplicativa e additiva) per le singole variabili e le loro interazioni.

Tabella 17 - Proseguimento degli studi per sesso e per titolo di studio del padre

Frequenze F_{ij}

	T ₁ basso			T ₂ medio-alto			T ₁ + T ₂		
	S1(M)	S2(F)	Tot.	S1	S2	Tot.	S1	S2	Tot.
C ₁ Continuo	1246	1111	2357	827	823	1647	2070	1934	4004
C ₂ Non continuo	726	874	1600	185	205	390	911	1079	1990
Totale	1972	1985	3957	1009	1028	2037	2981	3013	5994

Log-frequenze y_{ij}

C ₁ Continuo	7.128	7.013	7.070	6.714	6.713	6.713	6.921	6.863	6.892
C ₂ Non continuo	6.588	6.773	6.680	5.220	5.323	5.721	5.904	6.048	5.976
Totale	6.858	6.893	6.875	5.967	6.018	5.992	6.412	6.455	6.434

Per ottenere la tabella a tre dimensioni di tipo 2x2x2 nel nostro esempio si è proceduto ad una tabulazione incrociata delle variabili CONTINUAZIONE DEGLI STUDI x SESSO x TITOLO DI STUDIO DEL PADRE. Quest'ultima variabile è stata dicotomizzata in: titolo di studio BASSO (T₁) comprendente le categorie "nessun titolo" e "licenza elementare" e MEDIO-ALTO (T₂) comprendente le categorie "licenza media", "diploma di qualifica professionale", "diploma di maturità" e "laurea".

I valori osservati delle frequenze e delle log-frequenze sono riportati in Tabella 17.

Il modello completo saturo è espresso dalle seguenti equazioni nel caso, rispettivamente, di log-probabilità e di log-frequenze:

Tabelle 2x2x2:

Modello log-lineare additivo

Modello saturo:
 (log-probabilità) $V_{ijk} = \mu + \lambda_i^C + \lambda_j^S + \lambda_k^T + \lambda_{ij}^{CS} + \lambda_{ik}^{CT} + \lambda_{jk}^{ST} + \lambda_{ijk}^{CST}$
 (log-frequenze) $y_{ijk} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_k^T + \lambda_{ij}^{CS} + \lambda_{ik}^{CT} + \lambda_{jk}^{ST} + \lambda_{ijk}^{CST}$

vincoli: $\sum \lambda_i^C = \sum \lambda_j^S = \dots = \sum \lambda_{ij}^{CS} = \dots = \sum \lambda_{ijk}^{CST} = 0$

Come abbiamo visto nel caso 2x2, anche per le tabelle a tre dimensioni ogni valore della tabella è scomposto in un effetto generale, un effetto riga, un effetto colonna per ciascuna modalità della terza variabile oltre agli effetti di interazione di primo e di secondo ordine.

Secondo l'ottica generale dell'analisi della varianza, ciascun valore individuale di cella viene visto ora come somma di otto componenti additive: media generale, eccedenza di riga e di colonna per ciascuna modalità della terza variabile, interazione di primo e di secondo ordine. Esprimiamo pertanto in questi termini le varie componenti delle log-probabilità e delle log-frequenze di cella.

A questo fine risulta utile introdurre, per il caso delle log-frequenze, i simboli generali di ciascuna frequenza di cella e dei parziali marginali di riga, di colonna e del totale come risulta dal seguente schema:

	T ₁		Tot.	T ₂		Tot.	(T ₁ +T ₂)		
	S1	S2		S1	S2		S1	S2	Tot.
C ₁ continuo	y ₁₁₁	y ₁₂₁	y _{1.1}	y ₁₁₂	y ₁₂₂	y _{1.2}	y _{11.}	y _{12.}	y _{1..}
C ₂ non continuo	y ₂₁₁	y ₂₂₁	y _{2.1}	y ₂₁₂	y ₂₂₂	y _{2.2}	y _{21.}	y _{22.}	y _{2..}
Totale	y _{.11}	y _{.21}	y _{.1}	y _{.12}	y _{.22}	y _{.2}	y _{.1.}	y _{.2.}	y _{...}

Effetto medio

$$\Theta = y_{...} = \sum_i \sum_j \sum_k \frac{y_{ijk}}{IJK}$$

$$y_{i..} = \sum_j \sum_k \frac{y_{ijk}}{JK}$$

$$y_{ij.} = \sum_k \frac{y_{ijk}}{K}$$

$$y_{i.k} = \sum_j \frac{y_{ijk}}{J}$$

Date queste notazioni e dati i valori osservati delle frequenze di cella, così come risultano dalla Tabella 17, possiamo ricavare i valori degli otto parametri del modello saturo.

$$\begin{aligned} \Theta &= y_{...} &&= 6.434 \\ \lambda_i^C &= y_{i..} - y_{...} &&= 6.892 - 6.434 = 0.548 \quad (C_1 > C_2) \\ \lambda_j^S &= y_{.j.} - y_{...} &&= 6.412 - 6.434 = -0.022 \quad (S_1 < S_2) \\ \lambda_k^T &= y_{..k} - y_{...} &&= 6.875 - 6.434 = +0.441 \quad (T_1 > T_2) \\ \lambda_{ij}^{CS} &= y_{ij.} - y_{i..} - y_{.j.} + y_{...} &&= 6.921 - 6.892 - 6.412 + 6.434 = 0.051 \quad (C_1 S_1 > C_1 S_2) \\ \lambda_{ik}^{CT} &= y_{i.k} - y_{i..} - y_{..k} + y_{...} &&= 7.070 - 6.892 - 6.875 + 6.434 = -0.263 \quad (C_1 T_1 < C_1 T_2) \\ \lambda_{jk}^{ST} &= y_{.jk} - y_{.j.} - y_{..k} + y_{...} &&= 6.858 - 6.412 - 6.875 + 6.434 = 0.005 \quad (S_1 T_1 > S_1 T_2) \\ \lambda_{ijk}^{CST} &= y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...} &&= 7.128 - 6.921 - 7.070 - 6.858 + \\ &&&+ 6.892 + 6.412 + 6.875 - 6.434 = 0.024 \quad (C_1 S_1 T_1 > C_1 S_2 T_1, ecc.) \end{aligned}$$

Applicando i valori dei parametri calcolati all'equazione del modello saturo si ottengono i valori delle log-frequenze e delle frequenze come segue:

$$y_{111} = \Theta + \lambda_i^C + \lambda_j^S + \lambda_k^T + \lambda_{ij}^{CS} + \lambda_{ik}^{CT} + \lambda_{jk}^{ST} + \lambda_{ijk}^{CST}$$

$$= 6.434 + 0.458 - 0.022 + 0.441 + 0.051 - 0.263 + 0.005 + 0.024 = 7.128$$

$$F_{111} = \log y_{111} = \log (7.128) = 1246$$

e così via, per tutti i valori delle frequenze della tabella 2x2x2 che sono riportati in dettaglio nel prospetto che segue.

Il modello log-lineare saturo per tabelle 2x2x2

Modello moltiplicativo

	η	$\tau_{1..}^C$	$\tau_{.1.}^S$	$\tau_{..1}^T$	τ_{CS}	τ_{CT}	τ_{ST}	τ_{CST}	F_{ijk}
e ₁₁₁ =	622.65	1.581	0.979	1.555	1.052	0.769	1.004	1.025	1246
e ₁₂₁ =	622.65	1.581	1.021	1.555	0.951	0.769	0.996	1.025	1111
e ₂₁₁ =	622.65	0.633	0.979	1.555	0.951	1.300	1.004	0.976	726
e ₂₂₁ =	622.65	0.633	1.021	1.555	1.052	1.300	0.996	0.976	874
e ₁₁₂ =	622.65	1.581	0.979	0.643	1.052	1.300	0.996	0.976	824
e ₁₂₂ =	622.65	1.581	1.021	0.643	0.951	1.300	1.004	0.976	823
e ₂₁₂ =	622.65	0.633	0.979	0.643	0.951	0.769	0.996	1.025	185
e ₂₂₂ =	622.65	0.633	1.021	0.643	1.052	0.769	1.004	1.025	205

Modello additivo

	θ	λ^C	λ^S	λ^T	λ^{CS}	λ^{CT}	λ^{ST}	λ^{CST}	y_{ijk}	F_{ijk}
e ₁₁₁ =	6.434+0.458	-0.022	+0.441	+0.051	-0.263	+0.004	+0.025	7.128	1246	
e ₁₂₁ =	6.434+0.458	+0.022	+0.441	-0.051	-0.263	-0.004	+0.025	7.013	1111	
e ₂₁₁ =	6.434	-0.458	-0.022	+0.441	-0.051	+0.263	+0.004	-0.025	6.588	726
e ₂₂₁ =	6.434	-0.458	+0.022	+0.441	+0.051	+0.263	-0.004	-0.025	6.773	874
e ₁₁₂ =	6.434+0.458	-0.022	-0.441	+0.051	+0.263	-0.004	-0.025	6.714	824	
e ₁₂₂ =	6.434+0.458	+0.022	-0.441	-0.051	+0.263	+0.004	-0.025	6.713	823	
e ₂₁₂ =	6.434	-0.458	-0.022	-0.441	-0.051	-0.263	-0.004	+0.025	5.220	185
e ₂₂₂ =	6.434	-0.458	+0.022	-0.441	+0.051	-0.263	+0.004	+0.025	5.323	205

(1) λ^* 28.153 -1.324 27.131 3.105-16.165 0.235 1.509

(2) $\lambda^* = \frac{\lambda}{S.E.(x)}$

$$SE(\lambda) = \sqrt{\frac{\sum \sum \frac{1}{f_{ijk}}}{c^2}} = \sqrt{\frac{0.0169}{64}} = 0.016$$

Riportiamo nello schema di pagina seguente, anche per il modello moltiplicativo, le formule e il calcolo dei parametri in modelli saturi per tabelle 2x2x2.

Calcolo dei parametri in modelli saturi per tabelle 2x2x2

$$\eta = [F_{111} F_{121} F_{211} F_{221} F_{112} F_{122} F_{212} F_{222}]^{1/8} = \{1246 \cdot 1111 \cdot 726 \cdot 874 \cdot 824 \cdot 823 \cdot 185 \cdot 205\}^{1/8} = 623 \rightarrow \theta = 6.434$$

$$\tau_{1..}^C = [F_{111} F_{121} F_{112} F_{122}]^{1/4} / \eta = [1246 \cdot 1111 \cdot 824 \cdot 823]^{1/4} / 622.65 = 1581 \rightarrow \lambda_1^C = + 0.458$$

$$\tau_{.1.}^S = [F_{111} F_{121} F_{112} F_{122}]^{1/4} / \eta = [1246 \cdot 726 \cdot 824 \cdot 185]^{1/4} / 622.65 = 0.979 \rightarrow \lambda_1^S = - 0.022$$

$$\tau_{..1}^T = [F_{111} F_{121} F_{121} F_{221}]^{1/4} / \eta = [1246 \cdot 726 \cdot 1111 \cdot 874]^{1/4} / 622.65 = 1.555 \rightarrow \lambda_1^T = + 0.441$$

$$\tau_{11}^{CS} = \{ \tau_{CS/T1} \tau_{CS/T2} \}^{1/2} = \left\{ \left(\frac{F_{111}}{F_{211}} : \frac{F_{121}}{F_{221}} \right)^{1/4} \cdot \left(\frac{F_{112}}{F_{212}} : \frac{F_{122}}{F_{222}} \right)^{1/4} \right\}^{1/2} = \left\{ \left(\frac{1246}{726} : \frac{1111}{874} \right)^{1/4} \cdot \left(\frac{824}{185} : \frac{823}{205} \right)^{1/4} \right\}^{1/2} = \{1.350 \ 1.109\}^{1/8} = 1.052 \rightarrow \lambda^{CS} = 0.051$$

$$\tau_{1.1}^{CT} = \{ \tau_{CT/S1} \tau_{CT/S2} \}^{1/2} = \left\{ \left(\frac{F_{111}}{F_{211}} : \frac{F_{112}}{F_{212}} \right)^{1/4} \cdot \left(\frac{F_{121}}{F_{212}} : \frac{F_{122}}{F_{222}} \right)^{1/4} \right\}^{1/2} = \left\{ \left(\frac{1246}{726} : \frac{824}{185} \right)^{1/4} \cdot \left(\frac{111}{876} : \frac{823}{205} \right)^{1/4} \right\}^{1/2} = \{0.385 \ 0.317\}^{1/8} = 0.769 \rightarrow \lambda^{CT} = - 0.263$$

$$\tau_{.11}^{ST} = \{ \tau_{ST/C1} \tau_{ST/C2} \}^{1/2} = \left\{ \left(\frac{F_{111}}{F_{121}} : \frac{F_{112}}{F_{122}} \right)^{1/4} \cdot \left(\frac{F_{211}}{F_{221}} : \frac{F_{212}}{F_{222}} \right)^{1/4} \right\}^{1/2} = \left\{ \left(\frac{1246}{1111} : \frac{824}{823} \right)^{1/4} \cdot \left(\frac{726}{874} : \frac{185}{205} \right)^{1/4} \right\}^{1/8} = \{1.120 \ 0.921\}^{1/8} = 1.004 \rightarrow \lambda^{ST} = - 0.003$$

$$\tau_{111}^{CST} = \left\{ \left(\frac{F_{111}}{F_{211}} : \frac{F_{121}}{F_{221}} \right)^{T1} : \left(\frac{F_{112}}{F_{212}} : \frac{F_{122}}{F_{222}} \right)^{T2} \right\}^{1/8} = \left\{ \left(\frac{1246}{726} : \frac{1111}{874} \right) : \left(\frac{824}{185} : \frac{823}{205} \right) \right\}^{1/8} = \{(1.716 : 1.271) : (4.454 : 4.015)\}^{1/8} = \{1.350 : 1.109\}^{1/8} = 1.025 \rightarrow \lambda^{CST} = 0.0247$$

conditional odds ratio

E' bene ricordare che per il calcolo dei parametri τ in tabelle 2x2x2 valgono le seguenti relazioni:

$$\begin{aligned} \text{Effetti marginali } (\tau^{C..}, \tau^{S..}, \tau^{T..}) &= (\text{Prodotto degli Odds-product})^{1/8} \\ \text{es.: } \tau_{1..}^{C..} &= ((O/P)_{T1} \cdot (O/P)_{T2})^{1/8} \\ \text{Effetti parziali } (\tau^{CS.}, \tau^{C.T.}, \tau^{ST}) &= (\text{Prodotto degli Odds-ratio parziali})^{1/8} \\ \text{es.: } \tau_{11.}^{CS.} &= ((O/R)_{T1} \cdot (O/R)_{T2})^{1/8} \\ \text{Effetto interazione } (\tau^{CST}) &= (\text{Rapporto degli Odds-ratio parziali})^{1/8} \\ \text{es.: } \tau_{111}^{CST} &= ((O/R)_{T1} : (O/R)_{T2})^{1/8} \end{aligned}$$

Il significato di tali formule è abbastanza logico quando si ricorra ad alcune considerazioni pratiche:

a) Gli **effetti marginali** sono misure della 'asimmetria media combinata' di ogni singola variabile. Ad esempio il parametro $\tau_{1..}^{C..}$ rappresenta la misura dell'asimmetria fra coloro che continuano gli studi e coloro che non continuano, non più diretta, come nel caso 2x2, bensì mediata "attraverso" tutte le 4 celle-combinazione di Sesso x Titolo di studio. Il parametro $\tau^{C..}$ è pertanto la media geometrica degli odds product fra C_1 e C_2 , che misurano appunto le singole asimmetrie.

b) Gli **effetti parziali** rappresentano una misura dell'interazione (di primo ordine) esistente fra due variabili sotto controllo della terza. Ad esempio, il parametro $\tau_{11.}^{CS.}$ rappresenta l'effetto di interazione medio fra continuazione e sesso nei sottogruppi di studenti con genitori a Titolo di studio Basso e Medio-Alto. Nel nostro caso $\tau^{CS.} = 1.052$ indica che esiste un leggero effetto di interazione di primo ordine fra continuazione e sesso nel senso che i maschi tendono in generale a continuare gli studi più delle femmine; questa tendenza, tuttavia, è diversa per i due sottogruppi T_1 (Titolo di studio del padre basso) e T_2 (Titolo di studio del padre medio-alto). Infatti i due $\tau^{CS.}$ condizionati calcolati nei singoli sottogruppi T_1 e T_2 forniscono valori diversi, rispettivamente $\tau^{CS/T1} = 1.078$ e $\tau^{CS/T2} = 1.026$. Il $\tau^{CS.}$ parziale è la media geometrica dei due:

$$\tau^{CS.} = \left(\tau^{CS/T1} \cdot \tau^{CS/T2} \right)^{1/2} = \sqrt{1.078 \cdot 1.026} = 1.052$$

c) Gli **effetti interazione di secondo ordine** rappresentano le differenze fra i due effetti condizionati. Il nostro valore $\tau^{CST} = 1.025$ sta a significare appunto la (leggera) maggiore propensione a continuare gli studi dei maschi rispetto alle femmine riscontrata nei gruppi di studenti di livello sociale basso in rapporto a quelli di livello sociale più elevato.

Anche per il modello moltiplicativo il calcolo delle frequenze attese viene fatto attraverso i parametri del modello saturo come segue:

$$F_{111} = \eta \cdot \tau^C \cdot \tau^S \cdot \tau^T \cdot \tau^{CS} \cdot \tau^{CT} \cdot \tau^{ST} \cdot \tau^{CST}$$

da cui,

$$F_{111} = 623 \cdot 1.581 \cdot 0.979 \cdot 1.555 \cdot 1.061 \cdot 0.769 \cdot 1.003 \cdot 1.025 = 1.246$$

e così via per tutti i valori delle frequenze della tabella 2x2x2 riportati in dettaglio nel prospetto precedentemente illustrato.

2.2. - Il problema dei segni

Finora non abbiamo affrontato il problema dei segni dei singoli parametri-effetto che danno origine ai valori delle varie frequenze di cella attese F_{ijk} .

Pur derivando i segni direttamente dai vincoli posti dai vari modelli additivi e moltiplicativi già illustrati è bene rinforzare con un approccio concreto questo aspetto.

Caso di tabelle 2x2

Nel caso di modelli log-lineari per tabelle 2x2, i vincoli assegnati ai due

modelli additivi e moltiplicativi come già si è visto sono i seguenti:

Modelli additivi

$$\sum \lambda_{ij}^{CS} = 0$$

da cui:

$$\lambda_2^C = -\lambda_1^C$$

$$\lambda_2^S = -\lambda_1^S$$

$$\lambda_{11}^{CS} = \lambda_{22}^{CS} = -\lambda_{12}^{CS} = -\lambda_{21}^{CS}$$

Modelli moltiplicativi

$$\prod \tau_{ij}^{CS} = 1$$

da cui:

$$\tau_2^C = 1/\tau_1^C$$

$$\tau_2^S = 1/\tau_1^S$$

$$\tau_{11}^{CS} = \tau_{22}^{CS} = 1/\tau_{12}^{CS} = 1/\tau_{21}^{CS}$$

Dati questi vincoli, la matrice dei segni da assegnare ai vari modelli lambda nei modelli 2x2 è pertanto la seguente:

	(F _{ij})	C	S	CS
Continuano	F11	+	+	+
	F12	+	-	-
Non continuano	F21	-	+	-
	F22	-	-	+

L'attribuzione dei segni può essere impostata anche secondo la logica dei punteggi dummy (0,1) assegnati alle categorie delle variabili predittive di un modello di regressione.

Autori come Cox (1970) e Plackett (1974) utilizzano il "codice dummy" (dummy-coding) nel trattare i modelli log-lineari. Goodman assegna punteggi -1 e +1 alle categorie di ciascuna variabile dicotomica, procedura che va sotto il nome di "codice effetti" (effect-coding). I codici effetti sono appunto usati per generare i punteggi delle variabili indipendenti nell'equazione di previsione delle log-frequenze di celle attese.

I punteggi per i vari effetti lambda a una via si trovano assegnando a ciascuna cella della tabella 2x2 il punteggio -1 o +1 per la categoria di studenti (continuano/non continuano) associata a quella cella e

analogamente per assegnare i punteggi di cella per la variabile sesso.

La matrice dei segni si presenta nel modo seguente secondo le procedure "codice effetti" e "codice dummy" prima citate:

Codice effetti

	S ₁	S ₂	Tot.
C ₁	+1	-1	+1
C ₂	-1	+1	-1
Totale	+1	-1	

Codice dummy

	S ₁	S ₂	Tot.
C ₁	1	0	1
C ₂	0	1	0
Totale	1	0	

Il segno degli effetti-parametro λ_{ij}^{CS} , interazione tra continuazione degli studi e sesso, si trova moltiplicando i punteggi per gli effetti a una via (marginali) del proseguimento e del sesso associati con ciascuna singola cella.

Caso di tabelle 2x2x2

Nel caso di modelli log-lineari per tabelle 2x2x2, i vincoli assegnati ai parametri, come si è già visto, sono i seguenti:

Modelli additivi

$$\sum \lambda_{ijk}^{CST} = 0$$

$$\lambda_2^C = -\lambda_1^C$$

$$\lambda_2^S = -\lambda_1^S$$

$$\lambda_2^T = -\lambda_1^T$$

$$\lambda_{11}^{CS} = \lambda_{22}^{CS} = -\lambda_{12}^{CS} = -\lambda_{21}^{CS}$$

$$\lambda_{11}^{CT} = \lambda_{22}^{CT} = -\lambda_{12}^{CT} = -\lambda_{21}^{CT}$$

$$\lambda_{11}^{ST} = \lambda_{22}^{ST} = -\lambda_{12}^{ST} = -\lambda_{21}^{ST}$$

$$\lambda_{111}^{CST} = \lambda_{122}^{CST} = \lambda_{212}^{CST} = \lambda_{221}^{CST} =$$

$$= -\lambda_{112}^{CST} = -\lambda_{121}^{CST} = -\lambda_{211}^{CST} = -\lambda_{222}^{CST}$$

Modelli moltiplicativi

$$\prod \tau_{ijk}^{CST} = 1$$

$$\tau_2^C = 1/\tau_1^C$$

$$\tau_2^S = 1/\tau_1^S$$

$$\tau_2^T = 1/\tau_1^T$$

$$\tau_{11}^{CS} = \tau_{22}^{CS} = 1/\tau_{12}^{CS} = 1/\tau_{21}^{CS}$$

$$\tau_{11}^{CT} = \tau_{22}^{CT} = 1/\tau_{12}^{CT} = 1/\tau_{21}^{CT}$$

$$\tau_{11}^{ST} = \tau_{22}^{ST} = 1/\tau_{12}^{ST} = 1/\tau_{21}^{ST}$$

$$\tau_{111}^{CST} = \tau_{121}^{CST} = \tau_{212}^{CST} = \tau_{222}^{CST} =$$

$$= 1/\tau_{112}^{CST} = 1/\tau_{122}^{CST} = 1/\tau_{211}^{CST} = 1/\tau_{221}^{CST}$$

La matrice dei segni, per tabelle 2x2x2, è pertanto la seguente:

	(λ)			
	C	S	T	
F ₁₁₁	1	1	1	+ + + +
F ₁₁₂	1	1	2	+ - - -
F ₁₂₁	1	2	1	- + - -
F ₁₂₂	1	2	2	- - + +
F ₂₁₁	2	1	1	- - + -
F ₂₁₂	2	1	2	- + - +
F ₂₂₁	2	2	1	+ - - +
F ₂₂₂	2	2	2	+ + + -

2.3. - Il calcolo dei parametri nel caso di tabelle 2x2x2

Anche nel caso del modello log-lineare applicato a tabelle tridimensionali dicotomiche risulta possibile, e assai utile dal punto di vista euristico, procedere alla stima dei parametri-effetto τ e λ partendo dal calcolo degli odds, degli odds-ratio (di 1° ordine) e degli odds-ratio-ratio (di 2° ordine) secondo l'approccio di J. Davis (1974).

Per esigenze di completezza i calcoli sono stati sintetizzati nella Tabella 18 alla quale rinviamo i lettori interessati.

Nelle Tabelle n. 19 e n. 20 inoltre sono riportati in via analitica tutti i passaggi logici della procedura e i calcoli che permettono un confronto puntuale fra il modello degli odds e il modello log-lineare nelle sue varie forme. Viene anche riportato l'output del programma ECTA che analizza i dati della tavola 2x2x2 del nostro esempio relativo al proseguimento degli studi in provincia di Trento, distinto per sesso e per titolo di studio del padre.

Per quanto riguarda il commento al significato sostantivo dei risultati trovati a questo secondo livello, rinviamo il lettore al Capitolo 3.1 nel quale si affronterà con un'ottica esplicativa e con l'approccio logit l'interpretazione del nostro fenomeno proseguimento degli studi.

OUTPUT DEL PROGRAMMA ECTA PER TABELLE 2x2x2: MODELLO SATURO

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
4

NOTE
DISTRETTI IPOTESIB SESSOTITOLO 2x2x2
VAR1=IPOTESIB/ 1=CONTINUA 2=NON CONTINUA
VAR2=SESSO/1=MASCHIO 2=FEMMINA
VAR3=TITOLO/1=BASSO 2=ALTO

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
1 2 2 2

THE DIMENSION HAVE BEEN SET

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
4

NOTE

DATI

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
2

INPUT DATA READ
1246.00 726.00 1111.00 874.00 824.00 185.00 823.00 205.00

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
4

NOTE

(CST)

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
3 0 0 2 0 0 1 1 2 0 1 2 3

MARGINS FIT UNDER THE MODEL
1 2 3

AFTER ITERATION THE LARGEST DEVIATION IS .00000

PEARSON .00. AND LIKELIHOOD RATIO 0.00 CHI-SQUARE
0 DEGREES OF FREEDOM

THE ORIGINAL TABLE WITH THE FITTED VALUES SHOWN BENEATH IN EACH CELL
1ST. LEVEL OF VAR 3.

VAR 1	LEVELS OF VAR 2	
	1	2
1	1246.00	1111.00 I 2357.00
	1246.00	1111.00 I 2357.00
2	726.00	874.00 I 1600.00
	726.00	874.00 I 1600.00
-----		-----
	1972.00	1974.00 I 3957.00
	1972.00	1974.00 I 3957.00

2ND LEVEL OF VAR 3.		LEVELS OF VAR 2	
VAR 1	1	2	
1	824.00	823.00 I	1647.00
	824.00	823.00 I	1647.00
2	185.00	205.00 I	390.00
	185.00	205.00 I	390.00
	1009.00	1028.00 I	2037.00
	1009.00	1028.00 I	2037.00

MARGINAL TABLE		LEVELS OF VAR 1	
VAR 2	1	2	
1	1246.00	726.00 I	1972.00
	1111.00	874.00 I	1985.00
2	2357.00	1600.00 I	3957.00
	2357.00	1600.00 I	3957.00

2ND LEVEL OF VAR 3.		LEVELS OF VAR 1	
VAR 2	1	2	
1	824.00	185.00 I	1009.00
	823.00	205.00 I	1028.00
2	1647.00	390.00 I	2037.00
	1647.00	390.00 I	2037.00

THE ESTIMATED LAMBDA EFFECTS, THEIR STANDARD ERRORS, AND THE STANDARDIZED VALUES THESE THREE ARE FOLLOWED BY THE CORRESPONDING TAU PARAMETER OF THE MULTIPLICATIVE MODEL

VARIABLES OF ONLY TWO LEVELS WHERE THE SINGLE EFFECT SHOWN IS THE DIFFERENCE OF THE FIRST LEVEL, AND THE AVERAGE EFFECT

1	2	3	EFFECT	STD ERR	STDZD VAL	TAU PARAM
GRAND MEAN EFFECT			6.434			822.647
EFFECT FOR VARIABLES						
1			.458	.016	28.153	1.581
2			-.022	.016	-1.324	.979
3			.441	.016	27.131	1.655
1 2			.061	.018	3.105 **	1.052
1 3			-.263	.016	-16.165 ***	.769
2 3			.004	.016	.236	1.004
1 2 3			.025	.016	1.509	1.025

Tabella 18 - Calcolo degli odds nel caso di tabelle 2x2x2. Proseguimento degli studi per sesso e titolo di studio del padre.

ODDS PARZIALI

	CS		Odds	Odds (C)	CT		Odds	ST			
	S1	S2			T1	T2		T1	T2	Odds	
C1	2070	1934	1.070		2357	1647	1.431	S1	1972	1009	1.954
C2	911	1079	0.844	2.012	1600	390	4.102	S2	1985	1028	1.931
odds	2.272	1.792	1.268		1.473	4.223	0.349	0.993	0.985	1.012	
odds (s)		0.989	(O/R)				(O/R)			(O/R)	

ODDS CONDIZIONATI

	T1			Odds	(O/R)	T2			Odds	(O/R)
	S1	S2	Odds			S1	S2	Odds		
C1	1246	1111	1.121		824	823	1.001			
C2	726	874	0.831		185	205	0.902			
odds	1.716	1.271	1.350		4.454	4.015	1.109			
log-odds	0.540	0.240	0.300		1.494	1.390	0.104			

(O/R)_{II} = 1.350/1.109 = 1.217
log(O/R) = (0.300-0.104) = 0.196

e così in modo analogo:

	S1			Odds	(O/R)	S2			Odds	(O/R)
	T1	T2	Odds			T1	T2	Odds		
C1	1246	824	1.512		1111	823	1.350			
C2	726	185	3.924		874	205	4.263			
odds	1.716	4.454	0.385		1.271	4.015	0.317			

(O/R)_{II} = 0.385/0.317 = 1.217

	C1			Odds	(O/R)	C2			Odds	(O/R)
	S1	S2	Odds			S1	S2	Odds		
T1	1246	1111	1.121		726	874	0.831			
T2	824	823	1.001		185	205	0.902			
odds	1.512	1.350	1.120		3.924	4.263	0.921			

(O/R)_{II} = 1.120/0.920 = 1.217

Tabella 19 - Tabelle 2x2x2: Confronto fra modello degli odds e modello log-lineare

	T1			T2			T1 + T2			S1 + S2			C1 + C2			
	S1	S2	Tot.	S1	S2	Tot.	S1	S2	Tot.	T1	T2	Tot.	S1	T2	Tot.	
C1	1246	1111	2357	824	823	1647	2070	1934	4004	2357	1647	4004	S1	1972	1009	2981
C2	726	874	1600	185	205	390	911	1079	1990	1600	390	1990	S2	1985	1028	3013
Tot.	1972	1985	3957	1009	1028	2037	2981	3013	5994	3957	2037	5994	Tot.	3957	2037	5994

Simboli guida per le formule:

	T1			T2			T1+T2			S1+S2			C1+C2			
	S1	S2	Tot.	S1	S2	Tot.	S1	S2	Tot.	T1	T2	Tot.	T1	T2	Tot.	
C1	f ₁₁₁	f ₁₂₁	f _{1.1}	f ₁₁₂	f ₁₂₂	f _{1.2}	f _{1.1}	f _{1.2}	f _{1..}	f _{1.1}	f _{1.2}	f _{1..}	S1	f _{.11}	f _{.12}	f _{.1.}
C2	f ₂₁₁	f ₂₂₁	f _{2.1}	f ₂₁₂	f ₂₂₂	f _{2.2}	f _{2.1}	f _{2.2}	f _{2..}	f _{2.1}	f _{2.2}	f _{2..}	S2	f _{.21}	f _{.22}	f _{.2.}
Tot.	f _{.11}	f _{.21}	f _{.1.}	f _{.12}	f _{.22}	f _{.2.}	f _{.1.}	f _{.2.}	f _{...}	f _{.1}	f _{.2}	f _{...}	Tot.	f _{.1}	f _{.2}	f _{..}

	F _{ij}		M/F	B/M-A	Conditional odds	Partial odds CS/Tit.studio		
	C _i	S _j				T _k	I	II
	1	2			C/S,T	CS/T	CST	
	1246	726	1	1	1.716			
	1111	874	2	1	1.271	1.350		
	824	185	1	2	4.545		1.217	
	823	205	2	2	4.015	1.109		
Totale	4004	1990						
Marginal odds	(C):	4.004 : 1.990 = 2.012						
	(S):	2.981 : 3.013 = 0.989						
	(T):	3.957 : 2.037 = 1.943						
Conditional odds:	C/S1 T1	1.716						
	C/S2 T1	1.271						
	C/S1 T2	4.545						
	C/S2 T2	4.015						
Partial odds:	CS/T1	1.350						
(conditional odds ratio)								
Interazione di I ordine								
Partial odds ratio:	CST	1.217						
Interazione di II ordine								

(segue)

Tabella 19: (segue)

Calcolo dei parametri per i modelli log-lineari 2x2x2

Modello moltiplicativo

$$\eta = (1246 \cdot 1111 \cdot 726 \cdot 874 \cdot 824 \cdot 823 \cdot 185 \cdot 205)^{1/8} = 622.65$$

$$\tau_{1.}^C = \left[\left(\frac{F_{111}}{F_{211}} \frac{F_{121}}{F_{221}} \right) \cdot \left(\frac{F_{112}}{F_{212}} \frac{F_{122}}{F_{222}} \right) \right]^{1/8} \quad (\text{prodotto dei due odds products parziali CS/T})$$

$$\tau_{1.}^C = \left(\frac{1246}{726} \cdot \frac{1111}{874} \cdot \frac{824}{185} \cdot \frac{823}{205} \right)^{1/8} = 1.581$$

oppure formula abbreviata:

$$\tau_{1.}^C = (F_{111} F_{121} F_{112} F_{122})^{1/4} / \eta = (1246 \cdot 1111 \cdot 824 \cdot 823)^{1/4} / 622.65 = 1.581$$

$$\tau_{.1}^S = \left(\frac{F_{111}}{F_{121}} \frac{F_{211}}{F_{221}} \frac{F_{112}}{F_{122}} \frac{F_{212}}{F_{222}} \right)^{1/8} = \left(\frac{1246}{1111} \cdot \frac{726}{874} \cdot \frac{824}{823} \cdot \frac{185}{205} \right)^{1/8} = 0.979 \quad (\text{prodotto dei due odds product parziali SC/T})$$

oppure formula abbreviata:

$$\tau_{.1}^S = (F_{111} F_{211} F_{112} F_{212})^{1/4} / \eta = (1246 \cdot 726 \cdot 185)^{1/4} / 622.65 = 0.979$$

$$\tau_{.1}^T = \left(\frac{F_{111}}{F_{112}} \frac{F_{211}}{F_{212}} \frac{F_{121}}{F_{122}} \frac{F_{221}}{F_{222}} \right)^{1/8} = \left(\frac{1246}{824} \cdot \frac{726}{185} \cdot \frac{1111}{823} \cdot \frac{874}{205} \right)^{1/8} = 1.555 \quad (\text{prodotto dei due odds products parziali CT/S})$$

(segue)

Tabella 19: (segue)

oppure formula abbreviata:

$$\tau_1^T = (1246 \ 1111 \ 726 \ 874)^{1/4} / \eta = 1.555$$

$$\tau_{11}^{CS} = \left\{ \tau^{CS/T1} \tau^{CS/T2} \right\}^{1/2} = \left\{ \left(\frac{F_{111}}{F_{211}} : \frac{F_{121}}{F_{221}} \right)^{1/4} \cdot \left(\frac{F_{112}}{F_{212}} : \frac{F_{122}}{F_{222}} \right)^{1/4} \right\}^{1/2} = \left\{ \left(\frac{1246}{726} : \frac{1111}{874} \right)^{1/4} \cdot \left(\frac{824}{185} : \frac{823}{205} \right)^{1/4} \right\}^{1/2} = \{1.350 \ 1.109\}^{1/8} = 1.052 \ (\lambda^{CS} = 0.051)$$

$$\tau_{11}^{CT} = \left\{ \tau^{CT/S1} \tau^{CT/S2} \right\}^{1/2} = \left\{ \left(\frac{F_{111}}{F_{211}} : \frac{F_{112}}{F_{212}} \right)^{1/4} \cdot \left(\frac{F_{121}}{F_{221}} : \frac{F_{122}}{F_{222}} \right)^{1/4} \right\}^{1/2} = \left\{ \left(\frac{1246}{726} : \frac{824}{185} \right)^{1/4} \cdot \left(\frac{1111}{876} : \frac{823}{205} \right)^{1/4} \right\}^{1/2} = \{0.385 \ 0.317\}^{1/8} = 0.769 \ (\lambda^{CT} = -0.263)$$

$$\tau_{11}^{ST} = \left\{ \tau^{ST/C1} \tau^{ST/C2} \right\}^{1/2} = \left\{ \left(\frac{F_{111}}{F_{211}} : \frac{F_{112}}{F_{212}} \right)^{1/4} \cdot \left(\frac{F_{211}}{F_{221}} : \frac{F_{212}}{F_{222}} \right)^{1/4} \right\}^{1/2} = \left\{ \left(\frac{1246}{1111} : \frac{824}{823} \right)^{1/4} \cdot \left(\frac{726}{874} : \frac{185}{205} \right)^{1/4} \right\}^{1/2} = \{1.120 \ 0.921\}^{1/8} = 1.004 \ (\lambda^{ST} = 0.003)$$

$$\tau_{111}^{CST} = \left\{ \left(\frac{F_{111}}{F_{211}} : \frac{F_{121}}{F_{221}} \right) : \left(\frac{F_{112}}{F_{212}} : \frac{F_{122}}{F_{222}} \right) \right\}^{1/8} = 1.025$$

N.B.: $\tau^{CST} = (\text{odds ratio})^{1/8}$
 $1.025 = (1.217)^{1/8}$

Per il calcolo dei τ (2x2x2):

- Effetti marginali (C, S, T) = (Prodotto degli O/P)^{1/8} O/P = odds products
- Effetti parziali (CS, CT, ST) = (Prodotto degli O/R)^{1/8} O/R = odds ratio
- Effetti interazione (CST) = (Rapporto degli O/R)^{1/8}

(segue)

Tabella 19: (segue)

Modello additivo:

$$\Theta = y_{...} = \frac{\sum \sum \sum y_{ijk}}{8} = 6.434$$

$$\lambda_1^C = y_{1..} - y_{...} = 6.892 - 6.434 = +0.458$$

$$\lambda_j^S = y_{.j.} - y_{...} = 6.412 - 6.434 = -0.022$$

$$\lambda_k^T = y_{..k} - y_{...} = 6.875 - 6.434 = +0.041$$

$$\lambda_{ij}^{CS} = y_{ij.} - y_{i..} + y_{.j.} = 6.921 - 6.892 - 6.412 + 6.434 = 0.051$$

$$\lambda_{ik}^{CT} = y_{i.k} - y_{i..} - y_{.k.} + y_{...} = 7.070 - 6.892 - 6.875 + 6.434 = -0.263$$

$$\lambda_{jk}^{ST} = y_{.jk} - y_{.j.} - y_{..k} + y_{...} = 6.858 - 6.412 - 6.875 + 6.434 = +0.005$$

$$\lambda_{ijk}^{CST} = y_{ijk} - y_{ij.} - y_{i.k} - y_{.jk} + y_{i..} + y_{.j.} + y_{..k} - y_{...} =$$

$$= 7.128 - 6.921 - 7.070 - 6.858 + 6.892 + 6.412 + 6.875 - 6.434 = 0.024$$

Calcolo delle frequenze stimate (modello saturo)

Dal modello saturo moltiplicativo:

$$f_{ijk} = \eta \tau^C \tau^S \tau^T \tau^{CS} \tau^{CT} \tau^{ST} \tau^{CST}$$

$$f_{111} = 622.65 \cdot 1.581 \cdot 0.979 \cdot 1.555 \cdot 1.052 \cdot 0.769 \cdot 1.004 \cdot 1.025 = 1.246$$

Dal modello saturo additivo:

$$y_{ijk} = \Theta + \lambda^C + \lambda^S + \lambda^T + \lambda^{CS} + \lambda^{CT} + \lambda^{ST} + \lambda^{CST}$$

$$y_{111} = 6.434 + 0.458 - 0.022 + 0.441 + 0.051 - 0.263 + 0.005 + 0.024 = 7.128 \Rightarrow f_{111} = 1.246$$

Tabella 20 - Sintesi dei calcoli per impostare l'analisi con i vari modelli 2x2x2

Tabella originale (dati osservati)

C		S	T	ODDS CONDIZIONATI (C)			LOG-ODDS		
Continuano	Non continuano			I	II	III	I	II	III
1246	726	M +	BASSO -	1.716			0.540		
1111	874	F -	BASSO -	1.271	1.350		0.240	0.300	
824	185	M +	ALTO +	4.454		1.217	1.494		0.196
823	205	F -	ALTO +	4.015	1.109		1.390	0.104	

Modello log-lineare saturo:

Per ciascuna cella: $F_{ijk} = \eta \tau^C \tau^S \tau^T \tau^{CS} \tau^{CT} \tau^{ST} \tau^{CST}$

Simboli

Continuano	Non continuano	
F ₁₁₁	F ₂₁₁	$\eta = 622$
F ₁₂₁	F ₂₂₁	
F ₁₁₂	F ₂₁₂	
F ₁₂₂	F ₂₂₂	

	Indici dei vari τ								Segni (reciproci)							
	η	τ^C	τ^S	τ^T	τ^{CS}	τ^{CT}	τ^{ST}	τ^{CST}	τ^C	τ^S	τ^T	τ^{CS}	τ^{CT}	τ^{ST}	τ^{CST}	
CST																
F ₁₁₁ = 1246		1	1	1	11	11	11	111	+	+	+	+	+	+	+	
F ₁₂₁ = 1111		1	2	1	12	11	21	121	+	-	+	-	+	-	-	
F ₁₁₂ = 824		1	2	2	11	12	12	112	+	+	-	+	-	-	-	
F ₁₂₂ = 823		1	2	2	12	12	22	122	+	-	-	-	-	+	+	
F ₂₁₁ = 726		2	1	1	21	21	11	211	-	+	+	-	-	+	-	
F ₂₂₁ = 874		2	2	1	22	21	21	221	-	-	+	+	-	-	+	
F ₂₁₂ = 185		2	1	2	21	22	12	212	-	+	-	-	+	-	+	
F ₂₂₂ = 205		2	2	2	22	22	22	222	-	-	-	+	+	+	-	

(segue)

Tabella 20: (segue)

Modello Logit (effetti β sulla variabile dipendente C)

$$\Phi(\text{log-odds}) = \beta^C + \beta^{CS} + \beta^{CT} + \beta^{CST}$$

S T

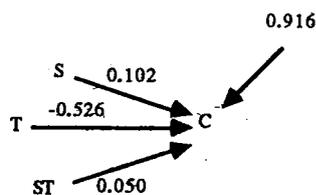
$$1 \ 1 \ \Phi_{11}^C = \beta^C + \beta_1^{CS} + \beta_1^{CT} + \beta_{11}^{CST} \quad 0.540 = 0.916 + 0.102 - 0.526 + 0.050$$

$$1 \ 2 \ \Phi_{12}^C = \beta^C + \beta_1^{CS} - \beta_2^{CT} - \beta_{12}^{CST} \quad 1.494 = 0.916 + 0.102 + 0.526 - 0.050$$

$$2 \ 1 \ \Phi_{21}^C = \beta^C - \beta_2^{CS} + \beta_1^{CT} - \beta_{21}^{CST} \quad 0.240 = 0.916 - 0.102 - 0.526 - 0.050$$

$$2 \ 2 \ \Phi_{22}^C = \beta^C - \beta_2^{CS} - \beta_2^{CT} + \beta_{22}^{CST} \quad 1.390 = 0.916 - 0.102 + 0.526 + 0.050$$

Modello ricorsivo:



Risposta:

- I maschi continuano più delle femmine
- I figli di genitori con titolo di studio alto continuano di più
- La maggiore propensione al proseguimento degli studi dei maschi rispetto alle femmine è più forte nelle classi basse.

N.B. I segni nel path diagram sono interpretabili con riferimento ai soggetti della cella (1,1)

Modelli non saturi per tabelle 2x2x2

Pur tenendo conto che in un'ottica tridimensionale è abbastanza semplice partire, secondo la procedura di tipo BACKWARD, dal modello saturo, vale la pena di analizzare i vari possibili modelli non saturi che si presentano in questo tipo di tabelle. Come mostra chiaramente la *Tally Chart* di Upton applicata ai nostri dati e riportata in Tabella n. 18, i possibili modelli gerarchici non saturi sono 19. Nel nostro esempio, gli unici modelli che, oltre al saturo, sono in grado di adattare in modo sistematicamente significativo le frequenze osservate sono i modelli [2] senza interazione (CS) (CT) (ST) e il modello [3] di indipendenza (CS) (CT). Il test di differenza tra frequenze osservate e attese produce un L^2 rispettivamente di 2.28 con 1 grado di libertà e di 3.65 con 2 gradi di libertà.

La Tabella n. 21 riporta analiticamente le frequenze delle 8 celle della tabella 2x2x2 generate da tutti i possibili modelli; dal saturo (CST) che adatta in modo perfetto i dati, al modello di equiprobabilità totale () che viceversa ipotizza un uguale numero atteso di soggetti per cella.

2.4. - Modelli per tabelle a quattro dimensioni con variabili dicotomiche (2x2x2x2)

Seguendo le notazioni relative ai modelli dicotomici (2x2x2) illustrate nel paragrafo precedente, la forma generale moltiplicativa del modello saturo (CSTZ) è la seguente:

$$F_{ijkl} = \eta \tau^C \tau^S \tau^T \tau^Z \tau^{CS} \tau^{CT} \tau^{CZ} \tau^{ST} \tau^{SZ} \tau^{TZ} \tau^{CST} \tau^{CSZ} \tau^{CTZ} \tau^{STZ} \tau^{CSTZ}$$

Il modello saturo esprime le sedici frequenze di cella attraverso un uguale numero di parametri.

Come si è già detto, il numero totale di modelli che si possono derivare da una tabella a 4 variabili dicotomiche è pari a 167 e appare pertanto

Tabella 21- Mappa dei possibili modelli gerarchici di relazione a tre variabili 2x2x2 (continuazione degli studi, sesso, titolo di studio del padre) (tally chart)

Modelli:		Parametri (λ)								Frequenze stimate										
		η	C	S	T	CS	CT	ST	CST	g.l.	L^2	p	F111	F121	F211	F221	F112	F122	F212	F222
1	Mod. saturo (CST)	6.434	.458	-.022	.441	.051	-.263	.004	.025	0	0		1246	1111	726	874	824	823	185	205
2	Mod. senza interazione (CS)(CT)(ST)	6.434	.459	-.032	.443	.063	-.264	.016		1	2.28	0.131	1234	1123	738	862	836	811	173	217
3	Mod. di indipendenza (CS)(CT)	6.434	.458	-.025	.443	.059	-.263			2	3.65	0.161	1219	1138	732	868	851	796	179	211
4	(CS)(ST)	6.503	.351	-.026	.332	.058		.003		2	268.38	0.000	1369	1274	603	711	701	660	308	368
5	(CT)(ST)	6.436	.457	-.006	.443		-.263	.003		2	22.23	0.000	1175	1182	797	803	816	831	193	197
6	(C)(ST)	6.505	.350	-.006	.332			.003		3	298.62	0.000	1317	1326	655	659	674	687	335	341
7	(S)(CT)	6.436	.457	-.005	.443		-.263			3	22.28	0.000	1172	1185	796	805	819	828	194	196
8	(T)(CS)	6.503	.351	-.025	.332	.059				3	278.55	0.000	1367	1277	601	712	703	657	310	367
9	(ST)	6.565		-.006	.332			.003		4	941.62	0.000	986	992	986	992	504	514	504	514
10	(C)	6.436	.457		.443		-.263			4	22.45	0.000	1178	1178	800	800	823	823	195	195
11	(CS)	6.557	.351	-.025		.059				4	864.99	0.000	1035	967	455	539	1035	967	455	539
12	(C)(S)(T)	6.505	.350	-.005	.332					4	298.40	0.000	1315	1329	653	660	677	684	336	340
13	(C)(S)	6.559	.350	-.005						5	892.43	0.000	996	1006	495	500	996	1006	495	500
14	(C)(T)	6.505	.350		.332					5	298.96	0.000	1322	1322	657	657	680	680	338	338
15	(S)(T)	6.565		-.005	.332					5	941.74	0.000	984	995	984	995	507	512	507	512
16	Mod. di equiprobabilità S, T (C)	6.559	.350							6	893.15	0.000	1001	1001	497	497	1001	1001	497	497
17	C, T (S)	6.619		-.005						6	1361.23	0.000	745	753	745	753	745	753	745	753
18	C, S (T)	6.565			.332					6	941.17	0.000	989	989	989	989	509	509	509	509
19	C, S, T ()	6.619								7	1360.48	0.000	749	749	749	749	749	749	749	749

C: C₁ Continuano, C₂ Non continuano; S: S₁ Maschi, S₂ Femmine; T: T₁ Titolo di studio basso, T₂ Titolo di studio medio-alto.

evidente come anche per l'impostazione di modelli relativamente semplici sia indispensabile il ricorso ad ipotesi teoricamente fondate. Anche in questo caso l'indagine esplorativa per la definizione del "miglior" modello potrà seguire le procedure FORWARD o BACKWARD in precedenza illustrate.

Le formule per il calcolo dei parametri del modello saturo sono le stesse viste per i modelli 2x2x2 con le ovvie modificazioni imposte dall'introduzione di una quarta variabile. Secondo l'ottica generale dell'analisi della varianza, ogni singola frequenza di cella risulta dalla sommatoria di 16 componenti additive ($i \lambda$ del modello log-lineare additivo) o dall'equivalente produttoria di 16 componenti moltiplicative ($i \tau$ del modello log-lineare moltiplicativo): media generale, eccedenze di riga e di colonna, interazioni di primo, di secondo e di terzo ordine.

La variabile aggiuntiva per ottenere il modello a 4 dimensioni è costituita, nel nostro esempio, dal livello socio-economico della zona di residenza dell'allunno (ZONA), dicotomizzato in "BASSO" e "MEDIO-ALTO".

La Tabella n. 22 riporta i dati della nostra ricerca sul proseguimento degli studi distinti per sesso, titolo di studio del padre e zona.

Vengono poi riportati i valori dei parametri del modello saturo così come sono prodotti e stampati dal programma ECTA di Goodman.

La Tabella n. 23 sintetizza i calcoli diretti effettuati sia con le formule abbreviate (Knoke, Burke, 1980, p. 16) sia con i valori degli odds di vario ordine (Davis, 1974).

Rispetto al modello a 3 variabili, l'inserimento nell'analisi della variabile ZONA consente la valutazione degli effetti diretti del livello socio-economico della zona di residenza dell'allunno sulla sua decisione di proseguire o meno gli studi (β_{CZ}), oltre che gli effetti di interazione fra titolo di studio e zona (β_{TZ}) e fra proseguimento, sesso e zona (β_{CSZ}). Gli altri effetti della variabile zona (β_{SZ} , β_{CTZ} , β_{STZ} e β_{CSTZ}) non risultano significativi.

Sulla base dell'analisi degli effetti-parametro del modello saturo a

Tabella 22 - (Modello 2x2x2x2) - Proseguimento degli studi per sesso, titolo di studio del padre e zona di residenza dell'allievo

(CSTZ)	Z1 Livello zona: basso T1 Titolo di studio: basso				Z1 Basso T2 Alto				Z2 Alto T1 Basso				Z2 Alto T2 Alto			
	S1	S2	Tot.	Odds	S1	S2	Tot.	Odds	S1	S2	Tot.	Odds	S1	S2	Tot.	Odds
C1	658	563	1221	1.169	204	186	390	1.097	531	511	1044	1.043	587	621	1208	0.945
C2	381	526	907	0.724	52	62	114	0.839	291	309	600	0.942	124	134	258	0.925
Tot.	1039	1089	2128	0.954	256	248	504	1.032	824	820	1644	1.005	711	755	1466	0.942
Odds	1.727	1.070	1.346	1.615	3.923	3.000	3.421	1.308	1.832	1.654	1.740	1.108	4.734	4.634	4.682	1.022
Log odds	0.546															
	(CS)				(CT)				(CZ)							
	S1	S2	Tot.	Odds	T1	T2	Tot.	Odds	Z1	Z2	Tot.	Odds				
C1	1982	1881	3863	1.054	2265	1598	3863	1.417	1611	2252	3863	0.715				
C2	848	1031	1879	0.823	1507	372	1879	4.051	1021	858	1879	1.190				
Tot.	2830	2912	5742	0.972	3772	1970	5742	1.915	2632	3110	5742	0.846				
Odds	2.337	1.824	2.056	1.281	(O/R) 1.503	4.296	2.056	0.350	1.578	2.625	2.056	0.601				
	(ST)				(SZ)				(TZ)							
	T1	T2	Tot.	Odds	Z1	Z2	Tot.	Odds	Z1	Z2	Tot.	Odds				
S1	1863	967	2830	1.927	1295	1535	2830	0.844	T1 2128	1644	3772	1.294				
S2	1909	1003	2912	1.903	1337	1575	2912	0.849	T2 504	1466	1970	0.344				
Tot.	3772	1970	5742	1.915	2632	3110	5742	0.846	2632	3110	5742	0.846				
Odds	0.976	0.964	0.972	1.012	0.969	0.975	0.972	0.994	4.222	1.121	1.915	3.766				

OUTPUT (PARZIALE) DEL PROGRAMMA ECTA PER TABELLE 2x2x2x2: MODELLO SATURO

THE ESTIMATED LAMBDA EFFECTS, THEIR STANDARD ERRORS, AND THE STANDARDIZED VALUES THESE THREE ARE FOLLOWED BY THE CORRESPONDING TAU PARAMETER OF THE MULTIPLICATIVE MODEL.

VARIABLES OF ONLY TWO LEVELS WHERE THE SINGLE EFFECT SHOWN IS THE DIFFERENCE OF THE FIRST LEVEL, AND THE AVERAGE EFFECT.

	1	2	3	4
GRAN MEAN EFFECT	5.681			
EFFECT FOR VARIABLES				
1	.455	.018	25.288 *	1.576
2	-.025	.018	-.1396	.975
3	.488	.018	27.125 *	1.629
4	-.175	.018	-9.707 *	.840
1 2	.054	.018	8.024 *	1.056
1 3	-.239	.018	-18.818 *	.787
1 4	-.070	.018	-8.880 *	.933
2 3	.002	.018	.114	1.002
2 4	-.008	.018	-.842	.994
3 4	.818	.018	17.418 *	1.368
1 2 3	.018	.018	1.012	1.018
1 2 4	.039	.018	2.166 *	1.040
1 3 4	.008	.018	.448	1.008
2 3 4	-.012	.018	-.692	.988
1 2 3 4	.008	.018	.449	1.008

Tabella 23 - Sintesi dei calcoli diretti dei parametri in modelli log-lineari 2x2x2x2

F _{ijk}	C1	C2	i	j	k	odds(C)	odds-ratio (CS)	odds-ratio-ratio (CST)	CSIZ
658	381	1	1	1	1	1.797	1.650	1.285	1.139
563	526	2	1	1	1.070	1.308			
204	52	1	2	1	3.923	1.108	1.084		
186	62	2	2	1	3.000				
533	291	1	1	2	1.832	1.028	1.084		
511	309	2	1	2	1.654				
587	124	1	2	2	4.734	1.028	1.084		
621	134	2	2	2	4.634				
π_{jki}						1448.4	2.392	1.339	
$(\pi_{jki})^{1/16}$						1.576	1.056	1.018	

(segue)

Tabella 23: (segue)

$$\eta = (\pi_{fijkl})^{1/16} = ((658 \cdot 563 \cdot 204 \cdot 186 \cdot 533 \cdot 511 \cdot 587 \cdot 621) \cdot (381 \cdot 526 \cdot 52 \cdot 62 \cdot 291 \cdot 309 \cdot 124 \cdot 134))^{1/16} = 278.99$$

Effetti marginali (C, S, T, Z) = (Prodotto degli Odds)^{1/16} es. $\tau_{1...}^C = (1.727 \cdot 1.070 \cdot 3.923 \cdot 3.000 \cdot 1.832 \cdot 1.654 \cdot 4.734 \cdot 4.634)^{1/16} = 1.576$
 Effetti interazione di I ordine (CS,CT,CZ,ST,SZ,TZ) = (Prodotto degli O/R)^{1/16} es. $\tau_{11..}^{CS} = (1.615 \cdot 1.308 \cdot 1.108 \cdot 1.022)^{1/16} = 1.056$
 Effetti interazione di II ordine (CST,CSZ,CTZ,STZ) = (Prodotto degli O/RR)^{1/16} es. $\tau_{111.}^{CST} = (1.235 \cdot 1.084)^{1/16} = 1.018$
 Effetti interazione di III ordine (C S T Z) = (Rapporto degli O/RR)^{1/16} es. $\tau_{1111}^{CSTZ} = (1.139)^{1/16} = 1.008$

dalle formule:

$$\tau_{1...}^C = (\pi_{f1...})^{1/8} / \eta = (658 \cdot 563 \cdot 204 \cdot 186 \cdot 533 \cdot 511 \cdot 587 \cdot 621)^{1/8} / 278.99 = 1.576$$

$$\tau_{1...}^S = (\pi_{f1...})^{1/8} / \eta = (658 \cdot 204 \cdot 533 \cdot 587 \cdot 381 \cdot 52 \cdot 291 \cdot 124)^{1/8} / 278.99 = 0.975$$

$$\tau_{11..}^{CS} = (\pi_{f11..})^{1/4} / \eta$$

Modello (2x2x2x2): Matrice degli indici e dei segni dei parametri c

Fijkl				τ^C	τ^S	τ^T	τ^Z	τ^{CS}	τ^{CT}	τ^{CZ}	τ^{ST}	τ^{SZ}	τ^{TZ}	τ^{CST}	τ^{CSZ}	τ^{CTZ}	τ^{STZ}	τ^{CSTZ}
1	1	1	1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
1	2	1	1	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-
1	1	2	1	+	+	-	+	+	-	+	-	+	-	+	-	-	-	-
1	2	2	1	+	-	-	+	-	-	+	-	-	+	-	-	-	+	+
1	1	1	2	+	+	+	-	+	+	-	+	-	+	-	-	-	-	-
1	2	1	2	+	-	+	-	-	+	-	-	+	-	+	-	-	+	+
1	1	2	2	+	+	-	-	+	-	-	-	+	-	-	+	+	+	+
1	2	2	2	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-
2	1	1	1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	2	1	1	+	-	+	+	-	+	+	-	+	-	-	+	-	-	-
2	1	2	1	+	+	-	+	+	+	+	-	+	-	+	-	-	-	-
2	2	2	1	+	-	-	+	-	-	+	-	-	+	-	-	-	+	+
2	1	1	2	+	+	+	-	+	+	-	+	-	+	-	-	-	-	-
2	2	1	2	+	-	+	-	-	+	-	-	+	-	+	-	-	+	+
2	1	2	2	+	+	-	-	+	-	-	-	+	-	-	+	+	+	+
2	2	2	2	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-

Tabella 24 - Analisi dei modelli log-lineari (2x2x2x2): procedura FORWARD

	Parametri τ																Significatività		
	η	C	S	T	Z	CS	CT	CZ	ST	SZ	TZ	CST	CSZ	CTZ	STZ	CSTZ	L ²	g.l.	p
1 (C S T Z)	278.99	1.576	.975	1.629	.840	1.056	.787	.933	1.002	.994	1.368	1.018	1.040	1.008	.988	1.008	0.0	0	1.000
2 (STZ)(C)	297.40	1.434	.991	1.475	.817				.998	1.005	1.393				.982		328.12	7	0.000
3 (STZ)(CS)	296.67	1.437	.970	1.475	.817	1.064			.998	1.005	1.393				.982		308.80	6	0.000
4 (STZ)(CT)	277.52	1.594	.991	1.650	.817		.769		.998	1.005	1.393				.982		52.65	6	0.000
5 (STZ)(CZ)	295.80	1.427	.991	1.475	.853			.881	.998	1.005	1.393				.982		246.95	6	0.000
6 (STZ)(CS)(CT)	276.67	1.598	.965	1.651	.817	1.068	.768		1.012	1.005	1.393				.982		31.97	5	0.000
7 (STZ)(CS)(CZ)	295.07	1.430	.970	1.475	.853	1.065		.880	.998	1.012	1.393				.982		227.47	5	0.000
9 (STZ)(CS)(CT)(CZ)	278.52	1.582	.965	1.634	.839	1.068	.784	.935	1.011	1.009	1.374				.983		11.72	4	0.190
10 (STZ)(CSZ)(CT)(ST)(TZ)	278.33	1.581	.968	1.633	.837	1.067	.784	.937	1.012	.989	1.374		1.048		.992		1.21	3	0.500
11 (STZ)(CST)(CZ)(SZ)(TZ)	278.79	1.580	.976	1.630	.839	1.053	.785	.935	.997	1.008	1.375	1.029			.984		8.86	3	0.310
12 (STZ)(CSZ)(CS)(ST)(SZ)	294.75	1.430	.976	1.475	.850	1.065		.882	.998	.995	1.393		1.050		.982		215.60	4	0.000
13 (STZ)(CSZ)(CST)(CTZ)	278.91	1.576	.974	1.629	.840	1.059	.787	.933	1.003	.991	1.369	1.016	1.044	1.008	.992		0.20	1	0.500
14 (TZ)(CS)(CT)(CZ)	278.71	1.581	.965	1.63	.839	1.064	.785	.935			1.375						14.40	7	0.044

quattro variabili, è possibile impostare alcuni modelli più parsimoniosi, la cui adattabilità ai dati della tabella empirica viene riportata in Tabella n. 24, dalla quale risultano tutti i parametri impliciti nei vari modelli e l'ammontare dei valori L^2 , g.l. e p.

Come si può vedere, gli unici modelli idonei a rappresentare in modo soddisfacente i dati della tabella risultano i seguenti:

Modello	L^2	g.l.	p.
(STZ)(CS)(CT)(CZ)	11.72	4	0.190
(STZ)(CSZ)(CT)(ST)(ZT)	1.21	3	0.500
(STZ)(CST)(CZ)(SZ)(TZ)	8.86	3	0.310
(STZ)(CSZ)(CST)(CTZ)	0.20	1	0.500
(TZ)(CS)(CT)(CZ)	14.40	7	0.044

Il fatto che i modelli log-lineari non distinguano fra variabili dipendenti e indipendenti costituisce un elemento di maggiore complessità dell'analisi, specie quando risulti chiaramente, dal contesto della ricerca, la definizione della variabile dipendente e la sequenza logica e temporale delle variabili. Come vedremo nel prossimo capitolo, il passaggio a modelli logit - e la conseguente definizione della variabile dipendente - semplifica l'analisi attraverso la riduzione del numero dei parametri β da calcolare e rende più chiara la definizione dell'intero quadro esplicativo.

2.5. - Modelli log-lineari con variabili politomiche

I modelli fin qui esaminati utilizzavano variabili dicotomiche. Qualora si vogliano trattare variabili politomiche il quadro dell'analisi presenta alcune rilevanti complicazioni non tanto per l'applicazione delle tecniche di stima che non vengono alterate dal numero di categorie delle variabili quanto dall'interpretazione dei risultati dell'analisi. Innanzitutto aumentano i parametri (τ , λ , β) da calcolare dal momento che aumentano i gradi di libertà della tabella multidimensionale e le relative combinazioni

("contrasti" (Upton, 1978, p. 105 e Taylor, 1983, p. 595 e segg.)) fra le diverse categorie delle variabili.

Se, ad esempio, anzichè utilizzare la variabile Titolo di studio del padre in forma dicotomica (BASSO/MEDIO-ALTO) si assume una tricotomia (BASSO/ MEDIO/ALTO) la tavola a due dimensioni (Continuazione x Titolo) comprende 6 celle (2x3) anzichè 4. I parametri da calcolare per il modello saturo diventano pertanto 6. Per il modello moltiplicativo essi si possono definire traendoli dal seguente elenco:

$$\eta, \tau, \tau_1^T, \tau_2^T, \tau_3^T, \tau_1^{CT}, \tau_2^{CT}, \tau_3^{CT}$$

tenendo conto che ciascuna tricotomia ha 2 gradi di libertà e richiede il calcolo di solo due effetti parametro. In questo caso si potranno scegliere due dei tre parametri τ_i^T e due dei tre parametri τ_i^{CT} .

Analogamente, in un modello 4x2x3 con una variabile dicotomica (SESSO (MASCHI/FEMMINE)) e con due variabili politomiche (continuazione degli studi (LICEI/ISTITUTI TECNICI/ISTITUTI PROFESSIONALI / NON CONTINUA) e titolo di studio del padre (BASSO/MEDIO/ALTO)) si dovranno calcolare 24 effetti parametro scelti fra l'insieme dei seguenti 32:

$$\eta, \tau_1^C, \tau_2^C, \tau_3^C, \tau_4^C, \tau_1^T, \tau_2^T, \tau_3^T, \tau_{11}^{CT}, \tau_{12}^{CT}, \tau_{13}^{CT}, \tau_{21}^{CT}, \tau_{22}^{CT}, \tau_{23}^{CT}, \tau_{31}^{CT}, \tau_{32}^{CT}, \tau_{33}^{CT}, \tau_{41}^{CT}, \tau_{42}^{CT}, \tau_{43}^{CT}$$

$$\tau_{111}^{CST}, \tau_{112}^{CST}, \tau_{113}^{CST}, \tau_{211}^{CST}, \tau_{212}^{CST}, \tau_{213}^{CST}, \tau_{311}^{CST}, \tau_{312}^{CST}, \tau_{313}^{CST}, \tau_{411}^{CST}, \tau_{412}^{CST}, \tau_{413}^{CST}$$

I 4 parametri τ_i^C potranno essere stimati in molti modi diversi a seconda della categoria che verrà scelta come base per il calcolo degli odds. Con quattro categorie si potranno stimare solo tre odds indipendenti (e i relativi parametri τ) e, più in generale, date K categorie, si dovranno calcolare K-1.

diversi parametri ¹.

Analogamente al modello D-SYSTEMS di Davis, che tratta le variabili dicotomiche creando singole *dummy* con valore (0, 1) per ciascuna delle K-1 categorie "non di base" della variabile originaria, nei modelli log-lineari si trasformano le politomie secondo i codici "dummy" o "effetti" (Taylor, 1983, p. 595).

In tal modo si creano due o più insiemi di punteggi che vengono usati per generare le variabili indipendenti nella regressione che predice i logaritmi delle frequenze di cella attese.

Nell'esempio che stiamo analizzando abbiamo 24 frequenze di cella e 24 parametri indipendenti da stimare:

A) 1 effetto media generale η

B) 6 effetti principali:

3 effetti principali C (tratti dall'insieme $\lambda_{C1}, \lambda_{C2}, \lambda_{C3}, \lambda_{C4}$);

2 effetti principali T (tratti dall'insieme $\lambda_{T1}, \lambda_{T2}, \lambda_{T3}$);

1 effetto principale S (tratto dall'insieme $\lambda_{S1}, \lambda_{S2}$)

C) 11 effetti interazione di primo ordine:

effetti di interazione (CS)(CT)(ST)

$$3 \cdot 1 + 3 \cdot 2 + 2 \cdot 1$$

D) 6 effetti di interazione di secondo ordine:

effetti di interazione (C S T)

$$3 \cdot 1 \cdot 2$$

La Tabella n. 25 riporta le frequenze stimate secondo i vari modelli testati nell'analisi. Come si può osservare i modelli che si adattano significativamente ai dati risultano essere il modello 1 (saturo: (CST)), il modello 2, che esclude l'interazione di secondo ordine ((CS)(CT)(ST)). Il modello 3 di indipendenza ((CS)(CT)) risulta invece più debole.

L'approccio FORWARD, analizzando con maggiore attenzione gli scarti fra frequenze di cella attese e osservate, consente una migliore

¹ Cfr. D.G. Taylor, 1983, pag. 596. Alcuni utili esempi di come usare i contrasti polinomiali per costruire una variabile dipendente vengono forniti da Duncan McRae (1978); Goodman (1971) e Haberman (1974b) e mostrano come usare i contrasti polinomiali per analizzare gli effetti di variabili indipendenti in una regressione logit.

articolazione del modello, ottenuta escludendo l'interazione di secondo ordine (CST) soltanto per il sottogruppo degli studenti con padre in possesso di titolo di studio medio e alto. Gli studenti con titolo di studio del padre basso, viceversa, presentano un fattore interazione (CST) significativo. In altre parole *nelle classi basse l'interazione fra abbandono degli studi e sesso è negativa nel senso che abbandonano di più le femmine; nelle classi alte tale interazione è invece positiva nel senso che abbandonano di più i maschi*. Il modello migliore e più parsimonioso che interpreta la continuazione degli studi in funzione del sesso e del titolo di studio paterno è pertanto il seguente: (T_{basso} S C) (C T).

I risultati analitici con tutti i parametri λ e τ del modello saturo prodotti dal programma ECTA sono riportati qui di seguito; nella Tabella n. 26, infine, vengono riassunti per tutti i modelli ipotizzati i parametri λ stimati, i valori L^2 , g.l. e p.

Tabella 25 - Modello log-lineare proseguimento degli studi per sesso e per titolo di studio del padre (4x2x3)

CST	Freq. osservate				e _{ijk} frequenze stimate secondo i vari modelli log-lineari															
	Fijk	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1 1 1	98	90	86	221	79	201	79	220	493	79	111	201	101	202	492	102	248	494	250
2	2 1 1	454	456	443	555	401	501	400	554	493	402	280	500	253	503	492	254	248	494	250
3	3 1 1	694	682	669	593	695	614	693	592	493	697	299	613	310	617	492	311	248	494	250
4	4 1 1	726	743	732	603	797	655	796	601	493	800	304	653	330	657	492	332	248	494	250
5	5 1 1	60	68	72	183	70	203	79	184	496	79	93	203	103	202	497	102	251	494	250
6	6 2 1	351	349	362	451	404	505	405	452	496	402	228	506	255	503	497	254	251	494	250
7	7 2 1	700	712	735	640	699	619	701	651	496	697	324	620	313	617	497	311	251	494	250
8	8 2 1	874	856	868	711	803	659	804	712	496	800	360	660	333	657	597	332	251	494	250
9	9 1 1	89	94	96	78	86	71	88	79	173	88	111	72	101	73	177	102	248	178	250
10	10 2 1	283	281	291	195	257	176	263	199	173	264	280	180	253	181	177	254	248	178	250
11	11 3 1	173	179	190	208	192	216	197	213	173	198	299	220	310	222	177	311	248	178	250
12	12 4 1	147	138	147	211	156	230	160	216	173	160	504	235	350	236	177	352	248	178	250
13	13 1 2	87	82	80	67	90	75	88	66	182	88	93	73	103	73	179	102	251	178	250
14	14 2 2	246	248	238	166	272	186	266	162	182	264	228	182	255	181	179	254	251	178	250
15	15 3 2	223	217	206	235	203	227	199	230	182	198	324	223	313	222	179	311	251	178	250
16	16 4 2	174	174	174	261	165	242	161	256	182	160	360	237	333	236	179	332	251	178	250
17	17 1 3	147	147	152	36	143	32	138	34	79	139	111	31	101	31	76	102	248	77	250
18	18 2 1	102	102	105	89	98	81	94	86	79	95	280	78	253	78	76	254	248	77	250
19	19 3 1	30	30	30	37	37	37	36	36	39	39	299	95	310	96	76	311	248	77	250
20	20 4 1	38	38	30	97	36	105	34	93	79	34	304	102	330	102	76	332	248	77	250
21	21 1 2	131	128	126	27	135	30	140	29	74	139	93	32	103	31	77	102	251	77	250
22	22 2	88	88	85	68	92	76	95	70	74	95	228	79	255	78	77	254	251	77	250
23	23 2	48	48	41	96	38	99	39	99	74	39	324	96	313	96	77	311	251	77	250
24	24 2	31	39	37	107	33	99	35	111	74	34	360	103	333	102	77	332	251	77	250
L ²	0	0	10.6	16.3	1324.6	52.4	1358.1	53.88	1324.86	1701.59	54.04	3973.23	1361.68	4030.60	1361.62	1703.12	4031.95	5835.85	1704.41	5338.89
g.l.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
p	0.00	0.101	0.037	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

OUTPUT DEL PROGRAMMA ECTA PER TABELLE 4x2x3: MODELLO SATURO

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
4

NOTE
DISTRETTI CON IPOTESIA SESSO TITOLO NON DICOTOMIZZATI 4x2x3
VARI=IPOTESIA/1=SERIE A 2=SERIE B 3=SERIE C 4=NON CONTINUA
VAR2=SESSO/1=MASCHIO 2=FEMMINA
VAR3=TITOLO/1=BASSO 2=MEDIO 3=ALTO

THE FOLLOWING NUMBER WERE READ FROM A CONTROL CARD
1 4 2 3

THE DIMENSION HAVE BEEN SET

THE FOLLOWING NUMBER WERE READ FROM A CONTROL CARD
2

INPUT DATA READ

98.00	454.00	694.00	726.00	60.00	351.00	700.00	874.00
89.00	283.00	173.00	147.00	87.00	246.00	223.00	174.00
147.00	102.00	30.00	38.00	131.00	88.00	48.00	31.00

THE FOLLOWING NUMBERS WERE READ FROM A CONTROL CARD
4

NOTE

(CST)

THE FOLLOWING NUMBER WERE READ FROM A CONTROL CARD
3 0 0 2 0 0 1 1 2 0 1 2 3

MARGINS FIT UNDER THE MODEL
1 2 3

AFTER ITERATION 2 THE LARGEST DEVIATION IS .00000

PEARSON .00, AND LIKELIHOOD RATIO 0.00 CHI-SQUARES
0 DEGREES OF FREEDOM

THE ORIGINAL TABLE WITH THE FITTED VALUES SHOWN BENEATH IN EACH CELL
1ST LEVEL OF VAR 3.

VAR 2	LEVELS OF VAR 1				I	
	1	2	3	4		
1	98.00	454.00	694.00	726.00	1972.00	
	98.00	454.00	694.00	726.00	1972.00	
2	60.00	351.00	700.00	874.00	1985.00	
	60.00	351.00	700.00	874.00	1985.00	
<hr/>						
	158.00	805.00	1394.00	1600.00	3957.00	
	158.00	805.00	1394.00	1600.00	3957.00	

2ND LEVEL OF VAR 3.

VAR 2	LEVELS OF VAR 1				I	
	1	2	3	4		
1	89.00	283.00	173.00	147.00	I	692.00
	89.00	283.00	173.00	147.00	I	692.00
2	87.00	246.00	223.00	174.00	I	730.00
	87.00	246.00	223.00	174.00	I	730.00
	176.00	529.00	396.00	321.00		1422.00
	176.00	529.00	396.00	321.00		1422.00

3RD LEVEL OF VAR 3.

VAR 2	LEVELS OF VAR 1				I	
	1	2	3	4		
1	147.00	102.00	30.00	38.00	I	317.00
	147.00	102.00	30.00	38.00	I	317.00
2	131.00	88.00	48.00	31.00	I	298.00
	131.00	88.00	48.00	31.00	I	298.00
	278.00	190.00	78.00	69.00	I	615.00
	278.00	190.00	78.00	69.00	I	615.00

MARGINAL TABLE

1ST LEVEL OF VAR 2.

VAR 3	LEVELS OF VAR 1				I	
	1	2	3	4		
1	98.00	454.00	694.00	726.00	I	1972.00
2	89.00	283.00	173.00	147.00	I	692.00
3	147.00	102.00	30.00	38.00	I	317.00
	334.00	839.00	897.00	911.00	I	2981.00

2ND LEVEL OF VAR 2.

VAR 3	LEVELS OF VAR 1				I	
	1	2	3	4		
1	60.00	351.00	700.00	874.00	I	1985.00
2	87.00	246.00	223.00	174.00	I	730.00
3	131.00	88.00	48.00	31.00	I	298.00
	278.00	685.00	971.00	1079.00	I	3013.00

THE ESTIMATED LAMBDA EFFECTS, THEIR STANDARD ERRORS, AND THE STANDARDIZED VALUES

THESE THREE ARE FOLLOWED BY THE CORRESPONDING TAU PARAMETER OF THE MULTIPLICATIVE MODEL

VARIABLES WHERE EACH LEVEL OF THE EFFECT IS THE DIFFERENCE BETWEEN THAT LEVEL OF THE VARIABLE, AND THE AVERAGE EFFECT

1 3

VARIABLES OF ONLY TWO LEVELS WHERE THE SINGLE EFFECT SHOWN IS THE DIFFERENCE OF THE FIRST LEVEL, AND THE AVERAGE EFFECT.

.2

	EFFECT	STD ERR	STDZD VAL	TAU PARM
GRAND MEAN EFFECT	5.052			156.280
EFFECT FOR VARIABLES				
1				
FOR LEVELS OF VAR 1				
1	-.468	.036	-12.971	.626
2	.320	.030	10.773	1.378
3	.103	.037	2.797	1.108
4	.045	.038	1.195	1.046
	.012	.020	.595	1.012
2				
3				
FOR LEVELS OF VAR 3				
1	.837	.025	33.935	2.310
2	.050	.026	1.911	1.052
3	-.888	.034	-25.762	.412
1 2				
FOR LEVELS OF VAR 1				
1	.093	.036	2.566	1.097
2	.079	.030	2.648	1.082
3	-.134	.037	-3.663	.874
4	-.037	.038	-.981	.963
2 3				
FOR LEVELS OF VAR 3				
1	.057	.025	2.315	1.059
2	-.045	.026	-1.695	.956
3	-.013	.034	-.364	.988

ESTIMATED LAMBDA EFFECTS, STANDARD ERRORS, AND STANDARDIZED VALUES FOR THE INTERACTION

VAR 3	LEVELS OF VAR 1			
	1	2	3	4
1	-1.081	-.220	.555	.746
	.051	.036	.041	.042
	-21.127	-6.128	13.608	17.899
2	-.156	.153	.078	-.072
	.050	.038	.045	.047
	-3.107	3.984	1.667	-1.532
3	1.237	.067	-.631	-.674
	.052	.050	.066	.068
	23.907	1.326	-9.529	-9.876

ESTIMATED TAU PARAMETERS FOR THE INTERACTION OF VARIABLES 1 2 3			
VAR 3	LEVELS OF VAR 1		
	1	2	3
1	.339	.803	1.742
2	.855	1.165	1.078
3	3.446	1.069	.532

ESTIMATED LAMBDA EFFECTS, STANDARD ERRORS, AND STANDARDIZED VALUES FOR THE INTERACTION OF VARIABLES 1 2 3			
VAR 3	LEVELS OF VAR 1		
	1	2	3
1	.083	-.019	-.061
2	.051	.038	.041
3	1.631	-.538	1.487

ESTIMATED TAU PARAMETERS FOR THE INTERACTION OF VARIABLES 1 2 3			
VAR 3	LEVELS OF VAR 1		
	1	2	3
1	.087	.981	1.063
2	.952	1.024	1.041
3	.966	.996	.905

Tabella 26 - Proseguimento degli studi per sesso e titolo di studio del padre (tabella 4x2x3)

C₁ Continuano serie A (liceo classico e scientifico)
 C₂ Continuano serie B (istituti tecnici e magistrali)
 C₃ Continuano serie C (istituti e scuole professionali)
 C₄ Non continuano

S₁ Maschi
 S₂ Femmine

T₁ Titolo di studio basso
 T₂ Titolo di studio medio
 T₃ Titolo di studio alto

Modelli		Parametri stimati (effetti λ)													g.l.	L ²	p						
		(η)	C ₂	C ₃	C ₄	S	T ₂	T ₃	C ₂ S	C ₃ S	C ₄ S	T ₂ S	T ₃ S	C ₂ T ₂				C ₃ T ₂	C ₄ T ₂	C ₂ T ₃	C ₃ T ₃	C ₄ T ₃	
1	Saturo (I)	(C S T)	5.052	.320	.103	.045	.000	.050	-.888	.000	.000	-.037	-.045	-.013	.153	.076	-.072	.067	-.631	-.674	0	.00	1.000
2	Senza interazione CST	(CS) (CT) (ST)	5.055	.317	.000	.000	.000	.046	-.886	.000	.000	-.117	-.027	-.019	.158	.074	-.072	.065	-.618	-.676	6	10.64	.101
3	Indipendenza	(CS) (CI)	5.056	.315	.000	.000	.000	.046	-.886	.000	.000	-.102			.155	.076	-.069	.064	-.616	-.674	8	16.33	.037
4		(CS) (ST)	5.145	.109	.000	.000	.000	-.062	-.900	.000	.000	-.102	-.027	.031							12	1324.61	.000
5		(CT) (ST)	5.059	.317	.000	.000	.000	.046	-.886				-.027	.031	.155	.076	-.069	.064	-.616	-.674	9	52.41	.000
6		(C) (ST)	5.148	.111	.000	.000	.000	-.062	-.900				-.027	.031							15	1358.18	.000
7		(S) (CT)	5.059	.317	.000	.000	.000	.046	-.886						.155	.076	-.069	.064	-.616	-.674	11	53.88	.000
8		(T) (CS)	5.145	.109	.000	.000	.000	-.062	-.900	.000	.000	-.102									14	1324.86	.000
9		(ST)	5.242				.000	-.062	-.900				-.027	.031							18	1701.59	.000
10		(CT)	5.059	.317	.000	.000		.046	-.886						.155	.076	-.069	.064	-.616	-.674	12	54.04	.000
11		(CS)	5.423	.109	.000	.000	.000			.000	.000	-.102									16	3973.23	.000
12		(C) (S) (T)	5.148	.111	.000	.000	.000	-.062	-.900												17	1361.68	.000
13		(C) (S)	5.427	.111	.000	.000	.000														19	4030.60	.000
14		(C) (T)	5.148	.111	.000	.000		-.062	-.900												18	1361.62	.000
15		(S) (T)	5.242				.000	-.062	-.900												20	1703.12	.000
16	Equiprobabilità	(C)	5.427	.111	.000	.000															20	4031.91	.000
17		(S)	5.520				.000														22	5835.95	.000
18		(T)	5.242																		21	1704.41	.000
19	Assenza di interazione	(η)	5.520																		23	5838.89	.000

(1) Il modello saturo (CST) comprende anche i sei effetti di interazione di II ordine tratti dai nove forniti dal programma ECTA:

- C₂ST₂ = 0.024
- C₂ST₃ = -0.004
- C₃ST₂ = 0.040
- C₃ST₃ = -0.100
- C₄ST₂ = -0.015
- C₄ST₃ = 0.139

CAPITOLO 3

I modelli logit

Finora abbiamo trattato il modello log-lineare generale, nel quale tutte le variabili sono considerate su un piano di parità, secondo un'ottica di configurazione di risposte la cui struttura deve essere determinata da una funzione moltiplicativa o additiva dell'intero insieme di variabili. La 'variabile criterio' da modellare attraverso i parametri-effetto era infatti la frequenza attesa di cella F_{ijk} . I modelli logit sono una versione particolare dei modelli log-lineari, in cui una variabile è assunta come dipendente rispetto alle altre. Essi pertanto risultano analoghi ai modelli ordinari di regressione lineare. Lo stesso Goodman (1972) a questo proposito parla di "approccio di regressione modificato".

La variabile criterio analizzata nei modelli logit è costituita dal logaritmo degli odds (o dagli odds) delle frequenze di cella attese per le due modalità dicotomiche della variabile dipendente. Nel nostro esempio il logaritmo del rapporto di probabilità (odds) continuano/non continuano rappresenta la variabile criterio del modello logit. Di regola il logit viene definito come segue (Knoke e Burke, 1980, p. 24):

$$\text{Logit} = \frac{1}{2} (\log\text{-odds})$$

Per confrontare il modello logit con il modello log-lineare generale consideriamo innanzitutto il caso a 3 variabili: proseguimento degli studi (C), sesso (S) e titolo di studio del padre (T).

Il proseguimento degli studi (C) rappresenta ora la variabile dipendente i cui odds sono funzione del sesso e del titolo di studio del padre.

Secondo l'espressione generale vista per il modello log-lineare saturo si ha:

$$F_{ijk} = \eta \tau_i^C \tau_j^S \tau_k^T \tau_{ij}^{CS} \tau_{ik}^{CT} \tau_{jk}^{ST} \tau_{ijk}^{CST}$$

Applicando questa espressione agli odds attesi sul proseguimento degli studi abbiamo (Knoke, Burke, 1980, p. 25):

$$\frac{F_{1jk}}{F_{2jk}} = \frac{\eta \tau_1^C \tau_j^S \tau_k^T \tau_{1j}^{CS} \tau_{1k}^{CT} \tau_{jk}^{ST} \tau_{1jk}^{CST}}{\eta \tau_2^C \tau_j^S \tau_k^T \tau_{2j}^{CS} \tau_{2k}^{CT} \tau_{jk}^{ST} \tau_{2jk}^{CST}}$$

da cui, semplificando, deriva che:

$$\frac{F_{1jk}}{F_{2jk}} = \frac{\tau_1^C \tau_{1j}^{CS} \tau_{1k}^{CT} \tau_{1jk}^{CST}}{\tau_2^C \tau_{2j}^{CS} \tau_{2k}^{CT} \tau_{2jk}^{CST}}$$

Ricordando i vincoli assunti dal modello moltiplicativo per variabili dicotomiche ($\prod \tau_{ijk}^{CST} = 1$) l'espressione si semplifica ulteriormente:

$$\frac{F_{1jk}}{F_{2jk}} = (\tau^C)^2 (\tau_j^{CS})^2 (\tau_k^{CT})^2 (\tau_{jk}^{CST})^2$$

Passando ai logaritmi si ha:

$$\text{Ln} \frac{F_{1jk}}{F_{2jk}} = 2 \text{Ln} (\tau^C) + 2 \text{Ln} (\tau_j^{CS}) + 2 \text{Ln} (\tau_k^{CT}) + 2 \text{Ln} (\tau_{jk}^{CST})$$

oppure:

$$\text{Ln} \frac{F_{1jk}}{F_{2jk}} = 2 \lambda^C + 2 \lambda_j^{CS} + 2 \lambda_k^{CT} + 2 \lambda_{jk}^{CST}$$

in cui

$$\lambda = \text{Log } \tau$$

Ponendo in questa espressione: $\beta = 2 \lambda$ si ottiene la notazione classica di Goodman (1972) che mostra chiaramente la relazione diretta esistente fra i parametri λ del modello log-lineare e quelli β del modello logit (Knoke, Burke, 1980, pp. 24-26):

$$\text{log-odds (C)} = \phi_{jk}^{(C)} = \beta^C + \beta_i^{CS} + \beta_k^{CT} + \beta_{jk}^{CST}$$

Il calcolo diretto dei log-odds utilizzabile nei modelli logit segue la stessa procedura indicata da Davis e già applicata in precedenza alle tabelle 2x2.

La Tabella n. 18 riporta il calcolo degli odds (semplici, parziali, condizionati) e dei relativi odds-ratio di primo e di secondo ordine. Mentre gli odds parziali e condizionati differiscono secondo l'impostazione data alla tabella multidimensionale, il valore dell'odds ratio di secondo ordine è unico. Il ricorso ai logaritmi degli odds permette l'applicazione immediata delle formule sopra riportate per i modelli logit.

Avendo calcolato i log-odds sulla variabile dipendente continuazione degli studi per le quattro combinazioni delle variabili indipendenti sesso e titolo di studio del padre, ed essendo già in possesso dei valori dei parametri β richiesti dal modello logit, ricavati secondo la formula $\beta = 2 \lambda$ dei valori λ forniti dal programma ECTA, siamo ora in grado di ottenere le quattro equazioni che corrispondono, spiegandoli, ai valori dei log-odds relativi alle quattro combinazioni di sesso x titolo di studio.

Dalla formula generale del modello logit applicato alle nostre variabili C, S, T si ha:

$$\phi^C = \beta^C + \beta^{CS} + \beta^{CT} + \beta^{CST}$$

Si hanno pertanto le seguenti quattro equazioni corrispondenti alle

combinazioni S x T:

S	T		
1	1	Maschi-titolo basso	$\phi_{11}^C = \beta^C + \beta_1^{CS} + \beta_1^{CT} + \beta_{11}^{CST}$
1	2	Maschi-titolo alto	$\phi_{12}^C = \beta^C + \beta_1^{CS} - \beta_2^{CT} - \beta_{12}^{CST}$
2	1	Femmine-titolo basso	$\phi_{21}^C = \beta^C - \beta_2^{CS} + \beta_1^{CT} - \beta_{21}^{CST}$
2	2	Femmine-titolo alto	$\phi_{22}^C = \beta^C - \beta_2^{CS} - \beta_2^{CT} + \beta_{22}^{CST}$

Sostituendo i valori dei log-odds e dei parametri β calcolati si ottengono (si veda la Tabella n. 20):

	(C)	(CS)	(CT)	(CST)
0.540 =	0.916	+0.102	- 0.526	+ 0.050
1.494 =	0.916	+ 0.102	+ 0.526	- 0.050
0.240 =	0.916	- 0.102	- 0.526	- 0.050
1.390 =	0.916	- 0.102	+ 0.526	+ 0.050

Per illustrare meglio le differenze tra modelli log-lineari e modelli logit analizziamo il caso di tabelle a 4 variabili (si vedano le tabelle 22 e 23). Supponiamo che gli odds sulla continuazione degli studi (C) dipendano dalle variabili sesso (S), titolo di studio del padre (T) e livello socio-economico della zona di residenza dell'allievo (Z).

L'equazione generale logit per questo modello è la seguente:

$$\phi_{jkl}^{(C)} = \beta^C + \beta_j^{CS} + \beta_k^{CT} + \beta_L^{CZ} + \beta_{jk}^{CST} + \beta_{jL}^{CSZ} + \beta_{kL}^{CTZ} + \beta_{jkl}^{CSTZ}$$

Tale equazione deriva direttamente dal modello log-lineare saturo (CSTZ) corrispondente. ϕ è il logaritmo degli odds attesi sulla continuazione degli studi e ciascun β è una media aritmetica dei log-odds per la continuazione degli studi calcolati attraverso tutti i livelli della particolare variabile indipendente o interazione indicata dalle lettere sottoscritte (Knoke e Burke, 1980, p. 26). Come per gli equivalenti

modelli log-lineari, la somma dei β per ciascun fattore che influenza (C) è pari a zero (ad es. $\beta_1^{CT} + \beta_2^{CT} = 0$).

Un importante aspetto del modello logit è che nell'equazione generale sopra riportata non compaiono gli effetti marginali e di interazione di primo e secondo ordine relativi alle variabili indipendenti, vale a dire (S), (T), (Z), (ST), (SZ), (TZ), (STZ). Pur non essendo esplicitamente compresi nell'equazione, questi valori marginali devono essere adattati nella stima delle frequenze attese sulle quali sono basati gli odds. La tabella marginale in cui tutte le variabili indipendenti interagiscono deve essere inclusa in ogni modello logit anche se il fattore non è statisticamente significativo.

Questa inclusione rappresenta la maggiore differenza nelle procedure di stima dei modelli logit rispetto ai modelli log-lineari generali (Knoke e Burke, 1980, p. 26).

Per la scelta e la stima del "miglior modello" logit di spiegazione risulta utile procedere nel modo seguente:

- a) definire un modello di base che comprenda tutte le variabili indipendenti e le rispettive interazioni (STZ) oltre che il controllo della distribuzione della variabile dipendente (C).
- b) Inserire su questo primo modello di base le varie ipotesi di relazioni tra variabili indipendenti e variabile dipendente. Questa procedura è analoga all'analisi di regressione multipla in cui le correlazioni fra le variabili indipendenti del modello sono tenute in considerazione anche se non appaiono in modo esplicito nell'equazione di regressione.

Nel caso del proseguimento degli studi in provincia di Trento, la sequenza FORWARD dei modelli ipotizzati potrebbe essere la seguente:

- (STZ)(C)
- (STZ)(CS)
- (STZ)(CT)
- (STZ)(CZ)
- (STZ)(CZ)(CT)
- (STZ)(CS)(CZ)
- (STZ)(CT)(CZ)
- (STZ)(CS)(CT)(CZ)

L'ultimo dei modelli sopra elencati adatta i dati in modo soddisfacente ($L^2 = 11.72$; g.l. = 4; $p = 0.190$) (cfr. Tabella n. 24) e si può assumere pertanto come base per un primo "parsimonioso" modello logit.

E' bene ricordare che soltanto la considerazione di tutti gli effetti presenti nel modello log-lineare saturo permette la definizione di un modello logit che dia conto di tutti gli elementi esplicativi della tabella multidimensionale.

c) Per ottenere i valori β del modello logit prescelto si ricorre ai valori τ del modello log-lineare corrispondente secondo la nota relazione

$$\beta = 2 \log \tau = 2\lambda$$

La Tabella n. 24 e l'output del programma ECTA visto in precedenza (pag.) riportano i dati per la costruzione dei vari modelli logit.

3.1. - Un modello logit per l'interpretazione del proseguimento degli studi in Provincia di Trento

3.1.1. - Il modello generale

Partendo dai parametri del modello log-lineare saturo della distribuzione multidimensionale CONTINUAZIONE x SESSO x TITOLO DI STUDIO DEL PADRE x LIVELLO SOCIO-ECONOMICO DELLA ZONA DI RESIDENZA è possibile costruire il modello logit generale:

$$\phi_{jkl}^{(C)} = \beta^C + \beta_j^{CS} + \beta_k^{CT} + \beta_l^{CZ} + \beta_{jk}^{CST} + \beta_{jl}^{CSZ} + \beta_{kl}^{CTZ} + \beta_{jkl}^{CSTZ}$$

Per il primo sottogruppo di soggetti (Maschi (1), Titolo del padre basso

(1), Zona bassa (1)) si ha:

$$\log \left(\frac{C_{1111}}{C_{2111}} \right) = \phi_{111}^{(C)} = \beta^C + \beta_1^{CS} + \beta_1^{CT} + \beta_1^{CZ} + \beta_{11}^{CST} + \beta_{11}^{CSZ} + \beta_{11}^{CTZ} + \beta_{111}^{CSTZ}$$

Sostituendo i valori β derivati dai λ del modello log-lineare saturo si ha:

$$\log \left(\frac{658}{381} \right) = 0.546 = 0.910 + 0.108 - 0.478 - 0.140 + 0.036 + 0.078 + 0.016 + 0.016$$

Esprimendo i risultati secondo l'usuale notazione della regressione multipla si possono così esprimere i vari effetti β sulla variabile dipendente C Continuazione degli studi:

Variabili indipendenti:

		(β)
Sesso (maschile)	SC	+ 0.109 *
Titolo (basso)	TC	- 0.479 *
Zona (bassa)	ZC	- 0.139 *
(Sesso x Titolo)	STC	+ 0.036
(Sesso x Zona)	SZC	+ 0.078 *
(Titolo x Zona)	TZC	+ 0.016
(Sesso x Titolo x Zona)	STZC	+ 0.016
Costante		+ 0.910 *

In altri termini, per entrare in un'ottica di spiegazione sociologica dei risultati ottenuti con il modello logit appena definito, possiamo affermare che in Provincia di Trento:

1. I maschi continuano più delle femmine ($\beta^{CS} = + 0.109$).
2. Il titolo di studio del padre basso diminuisce la propensione a continuare gli studi ($\beta^{CT} = - 0.479$).

3. Il basso livello socio-economico della zona di residenza diminuisce la propensione a continuare gli studi ($\beta^{CZ} = -0.139$).
4. Nella zona a più basso livello socio-economico i maschi continuano più delle femmine ($\beta^{CSZ} = +0.078$). In altri termini, la maggiore propensione agli studi post-obbligo dei maschi è maggiore nelle zone a più basso livello socio-economico.
5. La maggiore propensione agli studi post-obbligo dei maschi rispetto alle femmine è più forte nelle classi basse ($\beta^{CST} = +0.036$ non significativa) o, detto diversamente, nelle classi sociali più basse i maschi tendono a proseguire gli studi post-obbligo più delle femmine.
6. Sono pressochè nulli sia l'effetto interazione fra titolo di studio del padre e livello socio-economico della zona di residenza ($\beta^{CTZ} = +0.016$ n.s.) sia l'effetto interazione fra sesso, titolo di studio del padre e livello S.E. della zona di residenza ($\beta^{CSTZ} = +0.016$).

In alternativa al modello saturo è possibile definire un altro modello logit di interpretazione del nostro fenomeno di proseguimento degli studi in provincia di Trento partendo dai parametri del modello log-lineare:

$$(CS) (CT) (CZ) (STZ)$$

Tale modello come si è visto è più "parsimonioso" in quanto ipotizza l'assenza dei fattori di interazione di secondo e di terzo ordine fra la continuazione degli studi e le variabili indipendenti (sesso, titolo, zona). Tale modello logit avrà la seguente forma generale:

$$\phi_{jkl}^{(C)} = \beta^C + \beta_j^{CS} + \beta_k^{CT} + \beta_l^{TZ}$$

da cui:

$$\phi_{111}^{(C)} = \beta^C + \beta_1^{CS} + \beta_1^{CT} + \beta_1^{CZ}$$

Sostituendo i valori β derivati dal modello (9), ricavabili dai valori τ riportati in Tabella n. 24, si ha:

$$0.546 = 0.917 + 0.132 - 0.487 - 0.134 + \text{RES}$$

$$0.546 = 0.428 + 0.118$$

L'ammontare del residuo, pari a 0.118, corrisponde a quattro fattori di interazione ipotizzati nulli del modello (9).

Tali fattori infatti, dal modello saturo, risultano essere

$$\text{RES} = \beta^{CST} + \beta^{CSZ} + \beta^{CTZ} + \beta^{CSTZ} \quad \text{da cui}$$

$$0.118 = 0.036 + 0.078 + 0.016 + 0.016$$

Appare evidente, a questo punto, come la maggiore parsimonia vada a scapito della spiegazione che, viceversa, era consentita dal modello completo.

Che la maggior propensione agli studi post-obbligo dei maschi rispetto alle femmine sia superiore nelle zone a basso livello socio-economico è un'importante informazione sociologica. Altrettanto importante è l'informazione fornita dall'altro fattore di interazione fra sesso e titolo di studio del padre che ci dice come nelle classi basse siano le femmine le prime a rimanere a casa da scuola. Forse ancora più importante è il diverso ammontare numerico dei due effetti interazione: l'interazione continuazione x sesso (la maggiore propensione agli studi dei maschi rispetto alle femmine) è maggiore nelle zone più lontane dai centri urbani di quanto non lo sia per i gruppi sociali meno scolarizzati. In altre parole, la zona di residenza e quindi fattori locali e culturali giocano un peso maggiore di quanto non operino in provincia di Trento fattori esclusivamente sociali.

E' bene ricordare che le zone definite in questa prima versione dicotomica del modello come zone a livello socio-economico "basso" tendono a concentrare 3 diversi fattori che operano tutti nella stessa direzione relativamente alla propensione agli studi superiori dei figli:

- 1) nelle zone più periferiche prevale ancora il modello di preferenza sociale per il maschio che a parità di successo scolastico viene comunque privilegiato dalla famiglia (modello culturale);
- 2) nelle zone periferiche, l'assenza di sedi scolastiche superiori entro distanze territoriali contenute si riflette sfavorevolmente, e particolarmente, sul proseguimento delle femmine al cui spostamento (pendolarismo o trasferimento) i genitori pongono maggiori ostacoli;
- 3) le situazioni territorialmente marginali nella provincia di Trento spesso coincidono con ambienti sia agricoli sia a struttura turistico-alberghiera in continua espansione. Da qui le maggiori chances o attese occupazionali (anche stagionali) che privilegiano ruoli femminili.

3.1.2. - I modelli per i diversi indirizzi di studi

Vediamo ora di applicare i due modelli logit appena illustrati per l'interpretazione dei dati complessivi sul proseguimento degli studi, vale a dire il modello saturo (CSTZ) e il modello senza interazione (CS) (CT) (CZ) (TZ), ai dati relativi ad ogni singolo indirizzo scolastico scelto. La Tabella n. 28 riporta i dati su iscritti e non iscritti ai licei, agli istituti tecnici e agli istituti professionali in provincia di Trento nell'anno scolastico 1974-75 classificati sulla base delle variabili indipendenti sesso, titolo di studio del padre e livello socio-economico della zona di residenza. La Tabella n. 29 riporta i valori dei parametri β (si tratta dei valori calcolati dal programma ECTA) calcolati per i seguenti modelli logit:

modello generale saturo:

$$\phi = \beta^C + \beta^{CS} + \beta^{CT} + \beta^{CZ} + \beta^{CST} + \beta^{CSZ} + \beta^{CTZ} + \beta^{CSTZ}$$

modello senza interazione:

$$\phi = \beta^C + \beta^{CS} + \beta^{CT} + \beta^{CZ}$$

Il confronto tra i parametri del modello generale saturo che opera sul totale degli studenti e quelli tratti dai corrispondenti modelli applicati ai dati degli iscritti ai diversi indirizzi scolastici consente di articolare ulteriormente il quadro esplicativo:

- 1) (CS) - Se in generale, come si è visto, i maschi continuano gli studi più delle femmine, la relazione fra proseguimento degli studi e sesso evidenziata dai 3 sottomodelli distinti per indirizzo è la seguente:
 - a) Nei licei questa maggior propensione all'iscrizione dei maschi rispetto alle femmine è massima ($\beta^{CS} = + 0.166$).
 - b) Negli istituti tecnici si ritrova ancora in misura significativa questo tipo di tendenza ($\beta^{CS} = + 0.146$).
 - c) Negli istituti professionali, viceversa, questa tendenza si inverte: le femmine sono più orientate dei maschi verso questo tipo di indirizzo scolastico ($\beta^{CS} = - 0.082$).

Analisi degli effetti di interazione:

	Licei	Ist. tecnici	Ist. Prof.	Totale
CS	+ 0.166	+ 0.146	- 0.082	+ 0.109
CT	- 0.914	- 0.358	+ 0.272	- 0.479
CST	0.086	0.040	+ 0.058	0.036
CS	+ 0.166	+ 0.146	- 0.082	+ 0.109
CZ	- 0.172	- 0.074	+ 0.054	- 0.139
CSZ	+ 0.048	- 0.036	+ 0.116	0.078
CT	- 0.914	- 0.358	+ 0.272	- 0.479
CZ	- 0.172	- 0.074	+ 0.054	- 0.139
CTZ	+ 0.068	- 0.018	- 0.096	0.016
CSTZ	+ 0.048	- 0.040	+ 0.038	0.016

- 2) (CT) - *In generale il basso titolo di studio del padre diminuisce la propensione verso il proseguimento degli studi; questa tendenza risulta massima per i licei ($\beta^{CT} = - 0.914$); è presente, sia pure in misura attenuata, per gli istituti tecnici ($\beta^{CT} = -0.358$), mentre si inverte per gli istituti professionali verso i quali sono significativamente più orientati gli studenti provenienti dai gruppi sociali meno istruiti ($\beta^{CT} = + 0.272$).*
- 3) (CZ) - *La residenza in zone a basso livello socio-economico tende a diminuire la propensione a continuare gli studi. Questa tendenza appare più marcata qualora si consideri l'indirizzo di studi liceale ($\beta^{CZ} = - 0.172$) di quanto non avvenga negli altri indirizzi. Anche in questo caso l'indirizzo professionale muove, sia pure lievemente ed in modo non significativo, in direzione diversa ($\beta^{CZ} = + 0.054$).*
- 4) (CST) - *La maggiore propensione dei maschi delle classi basse a proseguire gli studi in misura superiore alle femmine, non significativa se riferita all'intera popolazione di studenti al termine della scuola dell'obbligo, è più marcata e raggiunge valori quasi significativi nel caso dei licei ($\beta^{CST} = + 0.086$).*
- 5) (CSZ) - *L'effetto interazione fra continuazione degli studi, sesso e zona presente nel modello generale riferito al totale degli studenti, assume i suoi valori più alti per il sottogruppo di studenti orientati verso gli istituti professionali ($\beta^{CSZ} = + 0.116$). In altri termini, nelle zone a livello socio-economico più basso i maschi tendono a frequentare gli istituti professionali in misura superiore delle femmine. Questa informazione aggiuntiva è particolarmente interessante e rinforza le osservazioni fatte nel precedente paragrafo circa la situazione progressiva di svantaggio scolastico delle femmine provenienti da situazioni in cui si cumulano elementi di marginalità sociali, culturali e territoriali.*

- 6) (CTZ) - *Analoghe interpretazioni si possono fare circa gli effetti di interazione (CTZ) emersi nel modello di analisi delle iscrizioni agli istituti professionali. Se la relazione di primo ordine (CT) rivela una maggiore propensione verso gli istituti professionali degli studenti provenienti dalle classi meno istruite, questa tendenza viene attenuata nelle zone a basso livello socio-economico ($\beta^{CTZ} = - 0.096$) in cui, come si è già osservato, esistono accanto alle scelte scolastiche di indirizzo professionale, altre importanti occasioni occupazionali o vincoli di tipo culturale che frenano, per motivi diversi, il proseguimento di studi anche se brevi.*

Tabella 27 - Modelli 2x2x2x2 per indirizzo scolastico: Proseguimento degli studi per sesso, titolo di studio del padre, livello socio-economico della zona. Dati sulle iscrizioni alle classi prime dei licei, istituti tecnici e istituti professionali (a.s. 1974/75)

	Z1 (zona: basso) T1 (titolo: basso)			Z1 (zona: basso) T2 (titolo: alto)			Z2 (zona: alto) T1 (titolo: basso)			Z2 (zona: alto) T2 (titolo:alto)			Totale			
	S1	S2	Tot.	S1	S2	Tot.	S1	S2	Tot.	S1	S2	Tot.	S1	S2	Tot.	
Licei																
Iscritti	C1	52	28	80	46	39	85	42	31	73	186	175	361	326	273	599
Non iscritti	C2	987	1061	2048	210	209	419	782	789	1571	525	580	1105	2504	2639	5143
Totale		1039	1089	2128	256	248	504	824	820	1644	711	755	1466	2830	2912	5742
Istituti tecnici																
Iscritti	C1	217	191	408	92	77	169	222	147	369	273	254	527	804	669	1473
Non iscritti	C2	822	898	1720	164	171	335	602	673	1275	438	501	939	2026	2243	4269
Totale		1039	1089	2128	256	248	504	824	820	1644	711	755	1466	2830	2912	5742
Istituti professionali																
Iscritti	C1	389	344	733	66	70	136	267	333	600	128	192	320	850	939	1789
Non iscritti	C2	650	745	1395	190	178	368	557	487	1044	583	563	1146	1980	1973	3953
Totale		1039	1089	2128	256	248	504	824	820	1644	711	755	1466	2830	2912	5742
Totale																
Continuano	C1	658	563	1221	204	186	390	531	511	1042	587	621	1208	1980	1881	3861
Non cont.	C2	381	526	907	52	62	114	293	309	602	124	134	258	850	1031	1881
Totale		1039	1089	2128	256	248	504	824	820	1644	711	755	1466	2830	2912	5742

N.B.: L'ammontare dei non iscritti rappresenta coloro che non risultano iscritti a quel particolare indirizzo scolastico e pertanto contiene sia gli iscritti agli altri due indirizzi sia coloro che non continuano gli studi.

Tabella 28 - Modelli logit per la spiegazione del proseguimento degli studi, secondo l'indirizzo di studi scelto

A. Modello logit generale: $\Phi = \beta^C + \beta^{CS} + \beta^{CT} + \beta^{CZ} + \beta^{CST} + \beta^{CSZ} + \beta^{CTZ} + \beta^{CSTZ}$

Parametri β

	C	CS	CT	CZ	CST	CSZ	CTZ	CSTZ	L2	gl	p	Res
Totale	.910**	.109**	-.479**	-.139**	.036	.078*	.016	.016	0	0		0
Licei	-2.272**	.166**	-.914**	-.172*	.086	.048	.068	.048	0	0		0
Ist. Tecnici	-.990**	.146**	-.358**	-.074*	.040	-.036	-.018	-.040	0	0		0
Ist. Profess.	-.874**	-.082*	-.272**	.054	.058	.116**	-.096**	.038	0	0		0

B. Modello logit senza interazione: $\Phi = \beta^C + \beta^{CS} + \beta^{CT} + \beta^{CZ}$

	C	CS	CT	CZ	L2	gl	p	Res
Totale	.916	.124	-.484	-.134	11.72	4	0.190	0.118
Licei	-2.242**	.114*	-.916**	-.180**	9.14	7	0.242	
Ist. Tecnici	-.996**	.142**	-.348**	-.084*	7.75	7	0.354	
Ist. Profess.	-.904**	-.052	.298**	.000	37.70	7	0.000	

C. Modello log-lineare saturo

Parametri τ

EFFETTI:		TOTALE	LICEI	IST.TECNICI	IST. PROFESS.
Marginali	C (continuazione degli studi)	** 1.576	0.321**	0.609**	0.646**
	S (sesso)	0.975	1.063*	1.026	0.971
	T (titolo di studio del padre)	** 1.629	1.029	1.358**	1.561**
	Z (zona di residenza)	** 0.840	0.769**	0.804**	0.831**
Parziali	CS (sesso (maschile))	** 1.056	1.086**	1.076**	0.960*
	CT (titolo (basso))	** 0.787	0.633**	0.836**	1.146**
	CZ (zona (bassa))	** 0.933	0.917**	0.964*	1.027
Interazione di I ordine	CST (sesso x titolo)	1.018	1.044	1.020	1.030
	CSZ (sesso x zona)	* 1.040	1.024	0.982	1.060**
	CTZ (titolo x zona)	1.008	1.035	0.991	0.953**
Interazione di II ordine	CSTZ (sesso x titolo x zona)	1.008	1.024	0.980	1.019
η		278.99	177.33	277.31	285.69

N.B.: * effetto significativo con $p < 0.05$; ** effetto significativo con $p < 0.01$

CAPITOLO 4

Modelli log-lineari ed analisi causale

Prima di concludere questa panoramica introduttiva ai concetti e alle procedure di base dell'analisi log-lineare vediamo, in quest'ultimo capitolo, le condizioni e i limiti dell'applicazione di modelli logit alla costruzione di modelli causali ricorsivi¹. Accanto ai numerosi punti di contatto tra le procedure dell'analisi causale classica, applicabile a variabili quantitative, e l'analisi causale log-lineare che opera viceversa su dati qualitativi, esistono alcune limitazioni rilevanti che restringono notevolmente l'applicazione e i vantaggi di quest'ultima tecnica.

Le analogie sono rappresentate essenzialmente dalla strutturazione delle ipotesi contenute nel modello in un diagramma causale sottoposto ad identiche regole di costruzione e dall'utilizzo dei coefficienti β provenienti da analoghi modelli di regressione multipla o modelli logit che misurano la forza e il segno degli effetti che legano fra loro le variabili del modello.

Tuttavia il campo di applicazione delle procedure log-lineari subisce due limitazioni fondamentali:

- 1) L'applicazione pratica dei modelli causali log-lineari richiede essenzialmente poche variabili dicotomiche. Non appena il numero delle variabili cresce o sono presenti anche poche politomie il modello tende a divenire rapidamente oscuro o ingestibile.
- 2) Non è possibile il calcolo degli effetti indiretti che, nella path analysis classica, legano le singole variabili indipendenti del modello alla variabile dipendente attraverso la moltiplicazione dei singoli coefficienti causali intermedi.

Malgrado queste limitazioni, l'applicazione dei modelli causali log-lineari può essere utile - con la cautela suggerita da Knoke e Burke - "ogni volta che un'ipotesi possa trarre vantaggio da sequenze causali unidirezionali tra le variabili" (Knoke e Burke, 1980, p. 42).

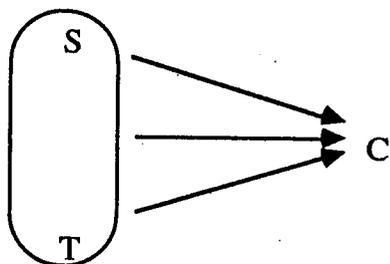
Il punto di partenza per la costruzione di un modello causale di relazioni tra variabili è un diagramma di effetti ricorsivi in cui, secondo le ben note regole dell'analisi causale, le variabili antecedenti vengono poste a sinistra

¹ Un'esposizione dettagliata della PATH ANALYSIS si può trovare in D. Heise (1975), H.B. Asher (1976), D. Duncan (1966, 1975) e L. Goodman (1979).

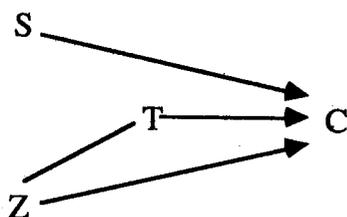
delle rispettive variabili conseguenti, le relazioni causali sono visualizzate da frecce unidirezionali, l'assenza di ordine causale è indicata da linee curve bidirezionali e l'assenza di relazione è indicata dall'assenza di frecce.

Riprendendo i modelli logit presentati nel paragrafo precedente traduciamo in diagrammi casuali le ipotesi ad essi sottostanti.

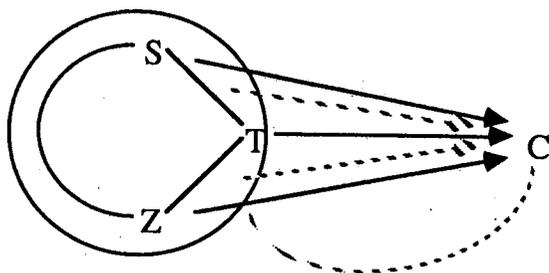
1° - Modello (CST)



2° - Modello (CS) (CT) (CZ) (TZ)



3° - Modello (CSTZ)



Le ipotesi sottostanti al primo modello sono le seguenti:

- 1) esiste una relazione tra sesso e continuazione degli studi (β^{CS})
- 2) esiste una relazione tra titolo di studio del padre e continuazione degli studi (β^{CT})
- 3) esiste un effetto interazione sesso x titolo nel senso che la relazione fra sesso e proseguimento degli studi è diversa per i due livelli di titoli di studio del padre (β^{CST})

Le ipotesi sottostanti al secondo modello sono le seguenti:

- a) esiste una relazione tra sesso e continuazione degli studi (β^{CS})
- b) esiste una relazione tra livello socio-economico della zona di residenza dell'allievo e proseguimento degli studi (β^{CZ})
- c) esiste una relazione tra titolo di studio del padre dell'allievo e proseguimento degli studi (β^{CT})
- d) esiste una relazione fra livello S.E. della zona di residenza dell'allievo e titolo di studio del padre (β^{TZ})
- e) non esistono effetti interazione delle tre variabili sesso, titolo e zona sul proseguimento degli studi.

Le ipotesi aggiuntive sottostanti del terzo modello sono le seguenti:

- f) esiste una relazione fra sesso e titolo di studio (β^{ST})
- g) esiste una interazione, di secondo ordine, fra sesso, titolo e proseguimento degli studi (β^{CST})
- h) esiste una interazione, di secondo ordine, fra zona, titolo e proseguimento degli studi (β^{CTZ})
- i) esiste una interazione di terzo ordine, fra sesso, titolo, zona e proseguimento degli studi (β^{CSTE}).

L'analisi causale dei dati della Tabella n. 23 differisce dalla costruzione del semplice modello logit in cui si fanno corrispondere, secondo l'ottica della regressione multipla, le variabili indipendenti (S, T, Z) sulla variabile dipendente (C).

I modelli causali devono tenere conto dell'ordine logico e temporale delle variabili, adattando una successione di modelli a varie tavole "compattate", appositamente costruite a partire dalla tavola globale.

Si procede con una serie di passi indipendenti, i cui risultati vengono alla fine riuniti ed interpretati.

1° passo: iniziando dalla parte sinistra del secondo diagramma, costruiamo la tabella 2x2 sesso x zona e adottiamo i modelli log-lineari per determinare se queste due variabili esogene sono correlate. Dato che il modello (S)(Z) ha un L^2 di 0.01 con 1 grado di libertà concludiamo che le due variabili sono indipendenti e nel diagramma non sono unite con la freccia curva bidirezionale che di regola unisce le variabili esogene dei modelli causali e riporta il coefficiente di correlazione lineare semplice. Il rapporto maschi/femmine è identico nelle due zone a livello socio-economico alto e basso.

2° passo: si costruisce la tabella a 3 vie formata dalle 2 variabili antecedenti S, Z sulla prima variabile dipendente nella sequenza del modello (T). Anche se è già stata verificata l'indipendenza fra le variabili sesso e zona, il modello logit che viene stimato in questa seconda fase richiede l'adattamento automatico della tabella marginale per tutti gli antecedenti causali della variabile Titolo. In tal modo l'analisi della struttura causale della sottotavola Sesso x Zona x Titolo deve includere la tabella marginale S x Z. Gli unici modelli da testare sono quindi quelli che comprendono la relazione del titolo di studio con i due antecedenti sesso e zona così come viene mostrato dalla Tabella n. 29.

Il migliore modello che interpreta i dati nella tabella a 3 vie è (SZ)(TZ) ($L^2 = 1.45$; g.l. = 2; $p = 0.485$).

3° passo: infine, si considera la continuazione degli studi (C) come variabile dipendente adattando la tavola marginale a 3 vie (Sesso x Zona x Titolo) nel processo di identificazione del modello logit maggiormente esplicativo delle frequenze osservate nella tabella globale a quattro

dimensioni.

In linea generale è buona regola mantenere controllato il set delle variabili indipendenti nella ricerca del miglior modello di spiegazione della C.

La Tabella n. 29 mostra vari modelli possibili secondo questa logica. Tra questi il miglior adattamento è fornito dal modello (STZ) (CS) (CT) (CZ).

L'analisi delle relazioni tra le tre variabili indipendenti compiuta nel 2° passo rende plausibile la definizione del modello più "parsimonioso", (TZ) (CS) (CT) (CZ) che, tuttavia, presenta un $L^2 = 14.40$ con g.l. = 7 e $p = 0.044$.

A questo punto si devono cumulare i risultati delle analisi compiute nei tre passi.

Il modello causale ricorsivo che meglio rappresenta i dati della Tabella n. 22 è la somma dei modelli delle tabulazioni incrociate successive a 2, a 3 e a 4 dimensioni. Tale modello adatta le tabelle marginali (S)(Z) - (SZ)(TZ) - (TZ)(CS)(CT)(CZ) ed ha $L^2 = (0.01 + 1.45 + 14.40) = 15.86$ con g.l. $(1 + 2 + 7) = 10$.

Un'alternativa a questo modello è: (S)(Z) (SZ)(TZ)(STZ)(CS) (CT)(CZ) che presenta $L^2 = (0.01 + 1.38 + 11.72) = 13.11$ con $(1 + 1 + 4) = 6$ g.l.

La stima dei parametri degli effetti causali del modello così impostato è data dai coefficienti beta che derivano dai modelli logit, così come è stato precedentemente illustrato, e che sono stati riportati nei rispettivi diagrammi causali.

Dato che il modello utilizza variabili dicotomiche, i valori beta di ciascuna relazione parziale possono essere interpretati come effetti delle variabili indipendenti sui log-odds delle variabili dipendenti. Si può vedere in tal modo come *dalle zone socio-economiche più elevate provengono studenti con più elevato titolo di studio del padre; inoltre dal diagramma causale emergono le relazioni significative che legano le tre variabili indipendenti sesso, titolo di studio del padre e zona di residenza al proseguimento degli studi. Infine, emerge il legame indiretto esistente fra*

Tabella 29 - Fasi della path analysis con modelli log lineari

I^a fase:

Dati da tabella a 2 vie (sesso x zona):

	Z1	Z2
S1	1295	1535
S2	1337	1575

Modelli (S) (Z)	L ²	g.l.	p
	0.01	1	>0.500**

II^a fase:

Dati da tabella a 3 vie (sesso x zona x titolo):

	Z1		Z2	
	T1	T2	T1	T2
S1	1039	256	824	711
S2	1089	248	820	755

Modelli:	L ²	g.l.	p
(SZ) (T)	514.37	3	0.000
(SZ) (TZ)	1.45	2	0.458**
(SZ) (TS)	514.33	2	0.000
(SZ) (TS) (TZ)	1.38	1	0.241**

III^a fase:

Dati da tabella a 4 vie (v. tabella n. 22) (continuazione x sesso x titolo x zona)

Modelli:	L ²	g.l.	p
(STZ) (C)	328.12	7	0.000
(STZ) (CS)	308.80	6	0.000
(STZ) (CT)	52.65	6	0.000
(STZ) (CZ)	246.95	6	0.000
(STZ) (CS) (CT)	31.97	5	0.000
(STZ) (CS) (CZ)	227.47	5	0.000
(STZ) (CT) (CZ)	32.26	5	0.000
(STZ) (CS) (CT) (CZ)	11.72	4	0.190 **
(TZ) (CS) (CT) (CZ)	14.40	7	0.044 *

IV. Risultati cumulativi:

Modello A:	I	II	III	Totale
	(S) (Z)	(SZ) (TZ)	(TZ) (CS) (CT) (CZ)	
L ²	0.01	1.45	14.40	15.86
g.l.	1	2	7	10
Modello B:	(S) (Z)	(SZ) (TS) (TZ)	(STZ) (CS) (CT) (CZ)	Totale
L ²	0.01	1.38	11.72	13.11
g.l.	1	1	9	6

zona e continuazione degli studi attraverso il titolo di studio del padre, nel senso che un elevato livello socio-economico della zona "produce" più elevati livelli sociali i quali, a loro volta inducono una maggiore propensione agli studi superiore dei figli.

Questa relazione indiretta è leggibile nel diagramma della contrapposizione fra effetto diretto Zona-Continuazione ($\beta^{ZC} = -0.134$) e effetti indiretti Zona-Titolo ($\beta^{ZT} = +0.635$) e Titolo-Continuazione ($\beta^{CT} = -0.484$)¹.

Tuttavia, ed è uno dei limiti più gravi dei modelli causali log-lineari, non è possibile, a differenza dei coefficienti dell'analisi causale quantitativa, moltiplicare i coefficienti path che collegano gli effetti indiretti ZT e TC per ottenere gli effetti diretti ZC.

Restano comunque valide le intuizioni sostantive suggerite dai segni e dall'ammontare numerico dei coefficienti.

Per concludere, il titolo di studio del padre ha un effetto diretto sulla continuazione degli studi superiore all'effetto diretto della zona, la quale influisce sulla continuazione degli studi anche in via indiretta, attraverso la variabile titolo di studio del padre.

Come già è stato osservato a proposito dei modelli logit, è possibile un confronto fra i valori numerici dei β per gli effetti diretti dato che entrambi sono espressi nella forma standardizzata di odds-ratio.

E' pertanto legittimo affermare che l'effetto diretto del titolo di studio è superiore all'effetto zona e che quest'ultimo è superiore all'effetto sesso.

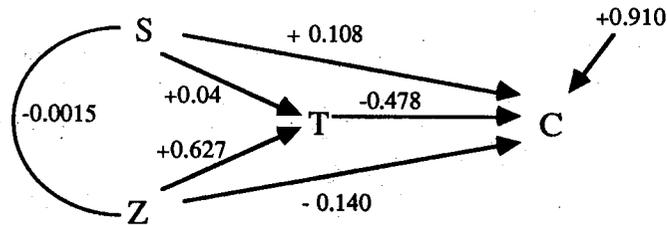
Finora abbiamo trascurato le interazioni di secondo e terzo ordine fra le variabili indipendenti (S, T, Z) e la continuazione degli studi (C). Da un'analisi attenta dei parametri del modello saturo (CSTZ) emerge l'importanza del fattore di interazione fra le variabili sesso, zona e continuazione degli studi ($\beta^{SZC} = +0.078$ significativo: $p < 0.01$).

Viceversa, gli altri effetti interazione (S x T x C) e (S x T x Z x C), pur presenti, non risultano significativi.

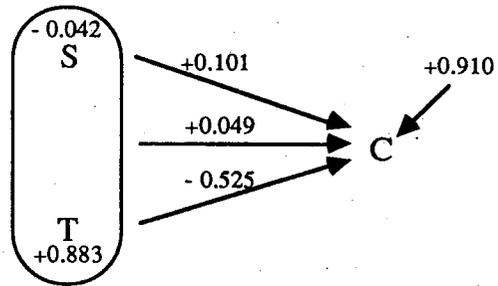
¹ I segni nel diagramma sono interpretabili con riferimento alla cella (1, 1, 1).

Per far emergere tali effetti di interazione e inserirli nel diagramma causale non vi è altra strada che assumere i parametri β del modello logit saturo, già riportati nel paragrafo precedente. Con tali integrazioni il diagramma causale del proseguimento degli studi risulta il seguente:

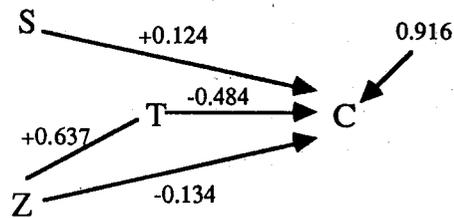
Modello (STZ) (CT) (CZ)



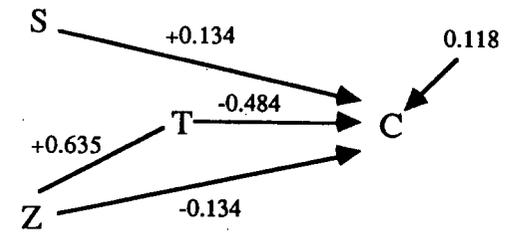
1° Modello: (CST)



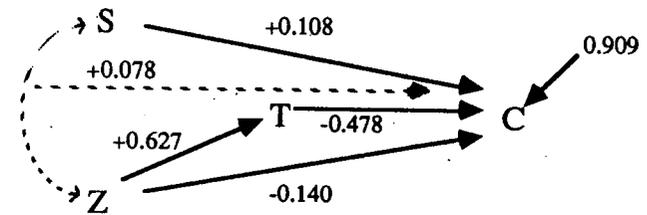
2° Modello: (CS) (CT) (CZ) (TZ)



3° Modello: (STZ) (CS) (CT) (CZ)



4° Modello: (CSTZ)



Conclusioni

Alla fine di questo lavoro, volutamente esposto in modo didattico, riteniamo utile richiamare l'attenzione su alcuni punti che rendono particolarmente produttivo il ricorso alla modellistica log-lineare.

1. Innanzitutto i modelli log-lineari, pur essendo di per sé validi da un punto di vista esplorativo, non possono prescindere da una adeguata elaborazione teorica che vede in essi un efficace ed immediato strumento di conferma. Con i modelli log-lineari è infatti possibile sottoporre a verifica empirica l'intero set di effetti causali di un fenomeno misurandone l'adattamento simultaneo ai dati empirici.
2. L'applicazione ottimale dei modelli log-lineari richiede un limitato numero di variabili e/o di categorie polinomiali. Come si è visto, la presenza di politomie o di un numero elevato di variabili rende difficilmente interpretabile il modello.
3. L'utilizzazione della procedura BACKWARD con l'analisi dei parametri del modello saturo è, dove possibile, un'utile base per l'analisi esplorativa. Viceversa, la procedura FORWARD si rivela più efficace in un'ottica confirmatoria.
4. I modelli logit, che operano seguendo la sintassi della regressione multipla con la definizione di una variabile dipendente, bene si adattano ai canoni della spiegazione sociologica. Attraverso la procedura logit, come si è visto, è possibile impostare un modello simile alla regressione multipla con variabili di tipo categoriale, colmando una notevole lacuna nell'analisi statistica di dati sociologici.
5. I modelli causali log-lineari, pur utilizzabili in modo più restrittivo rispetto ai corrispondenti modelli di path analysis per variabili quantitative, sono utili in quanto forniscono preziose indicazioni teoriche sia sulla direzione delle relazioni causali (i segni) sia sull'ordine di grandezza dei parametri beta.

6. Infine un motivo di cautela nell'applicazione dei modelli log-lineari è rappresentato dal fatto che, come osserva saggiamente G. Gangemi (1980), la verifica della *bontà di adattamento* è fondata, di fatto, sul test Chi quadrato che, com'è noto, risente della numerosità del campione perdendo rapidamente di efficacia all'aumentare di esso.

Pur con queste cautele e limiti riteniamo che questa procedura di analisi sia pienamente degna di attenzione e di applicazione nella ricerca sociologica empirica.

Bibliografia

Abdelbasit, K.M. & Plackett, R.L.

1981 Experimental design for categorized data, in *Int. Statist. Rev.*, 49, 111-126.

Affi, A.A. & Clark, V.

1984 *Computer-aided Multivariate Analysis*, Belmont, California. Lifetime Learning Publications.

Agresti, A.

1976 The effect of category choice on some ordinal measures of association, in *J. Amer. Statist. Assoc.*, 71, 49-55.

1977 Considerations in measuring partial association for ordinal categorical data, in *J. Amer. Statist. Assoc.*, 72, 37-45.

1980 Generalized odds ratios for ordinal data, in *Biometrics*, 36, 59-67.

1981 A hierarchical system of interaction measures for multidimensional contingency tables, in *J. Roy. Statist. Soc.*, B, 43, 293-301.

1981 Measures of nominal-order association, in *J. Amer. Statist. Assoc.*, 76, 254-529.

1983 A survey of strategies for modelling cross-classifications having ordinal variables, in *J. Amer. Statist. Assoc.*, 78, 184-198.

Agresti, A. & Kezouch, A.

1983 Association models for multi-dimensional cross-classifications of ordinal variables, in *Commun. Statist. Theor. Meth.*, 12, 1261-1276.

Aickin, M.

1982 Serial P-values, in *J. Statistical Planning and Inference*, 7, 243-256.

Aitkin, M.A.

1979 A simultaneous test procedure for contingency table models, in *Appl. Statist.*, 28, 233-242.

1980 A note on the selection of log-linear models, in *Biometrics*, 36, 173-178.

Albert, J.H. & Gupta, A.K.

1982 Mixtures of Dirichlet distributions and estimation in contingency tables, in *Ann. Statist.*, 10, 1261-1268.

1983 Estimation in contingency tables using prior information, *J. Roy. Statist. Soc.*, B, 45, 60-69.

Albrecht, P.

1980 On the correct use of the chi-square goodness-of-fit test, in *Scand. Act. J.*, 149-160.

Allison, P.D.

1980 Analyzing collapsed contingency tables without actually collapsing, in *Amer. Soc. Rev.*, 45, 123-130.

- Altham, P.M.E.
1976 Discrete variable analysis for individuals grouped into families, in *Biometrika*, 63, 263-269.
1979 Detecting relationship between categorical variables observed over time: a problem of deflating a chi-square statistic, in *Appl. Statist.*, 28, 115-125.
- Ammassari, P.
1986 L'analisi causale nella ricerca sociale. In: G. Amendola, A. Ardigò (a cura di), *Ricerca sociologica, Informatica e Società Italiana*. Milano, F. Angeli, 1986, pp. 59-102.
- Anderberg, M.R.
1973 *Cluster Analysis for Applications*, New York, Academic Press.
- Andersen, E.B.
1980 *Discrete statistical models with social science applications*, New York, North Holland.
- Anderson, R.J.
1980 Catania for multidimensional contingency tables: nominal-scale response, in *Comm. Stat. Theor. Meth.*, A9, 1191-1206.
- Andrews, D.F., Gnanadesikan, R. & Warner J.L.
1973 Methods for assessing multivariate normality, in Krishnaiah, P.R., *Multivariate Analysis III*, New York, Academic Press.
- Andrews, F.M., Morgan, J.N. & Sonquist, J.A.
1967 *Multiple classification analysis*, Ann Arbor: Institute for Social Research, Univ. of Michigan.
- Andrich, D.
1979 A model for contingency tables having ordered response classification, in *Biometrics*, 35, 403-415.
- Ansell, J., Bendell, A. & Humble, S.
1982 On the small estimation of the log odds ratio in 2x2 contingency tables with one set of fixed margins, in *J. Stat. Comp. Simul.*, 14, 305-320.
- Arminger, G.
1984 Analysis of qualitative individual data and of latent class models with generalized linear models, in Nijkamp, P.J. (ed), *Measuring the Unmeasurable*, Amsterdam, North Holland.
- Barcikowski, R.S. & Robey, R.R.
1984 Decisions in single group repeated measures analysis: statistical tests and three computer packages, in *American Statistician*, 38:2, 148-150.

- Barnard, G.A.
1980 Note on an especially simple partitioning of chi-square from questionnaires, in *J. Amer. Psy. Assoc.*, 58.
1981 The conditional approach to robustness, in Csorgo, M., Dawson, D.A., Rao, J.N.K. & Saleh, A.K.Md. (eds.), *Statistics and related topics*, New York, North Holland.
1982 Conditionality versus similarity in the analysis of 2x2 tables, in Kallinapur, G., Krishnaiah, P.R. & Ghosh, J.K. (eds.), *Statistics and Probability: Essays in Honor of C.R. Rao*, New York, North Holland.
- Barnard, R.M.
1973 *Field-dependent Independence and selected Motor Abilities*, Ph. D. diss., School of Education of New York University.
- Bartholomew, D.J.
1980 Factor analysis for categorical data, in *J. Roy. Statist. Soc.*, B 42, 293-321.
- Bedrick, E.J.
1983 Adjusted chi-square tests for cross-classified tables of survey data, in *Biometrika*, 70, 591-595.
- Benedetti, J.K. & Brown, M.B.
1978 Strategies for the selection of log-linear models, in *Biometrics*, 34, 680-686.
- Berk, R.A.
1983 Applications of general linear model to survey data, in P. Rossi, J.S. Smith & Anderson, A.B., *Handbook of survey research*, New York, McMillan.
- Berkson, J.
1978 In dispraise of the exact test, in *J. Statist. Plann. Inf.*, 2, 27-42.
- Bhapkar, V.P. & Koch Gary, G.
1968 On the hypothesis of no interaction in multidimensional contingency tables, in *Technometrics*, 10, 107-124.
- Bishop, Y.M.M., Fienberg, S. & Holland, P.
1975 *Discrete Multivariate Analysis: Theory and Practice*, Cambridge: MIT Press.
- Bjornstad, J.F.
1982 Comparison of dependent two-way contingency tables, in *Comm. Statist. Theor. Meth.*, 11, 673-686.
- Blyth, C.R.
1972 On Simpson's paradox and the sure-thing principle, in *Journal of the American Statistical Association*, 67, 364-366.

- Bock, R.D.
1975 *Multivariate Statistical Methods in Behavioral Research*, New York, McGraw-Hill.
- Breiger, R.L.
1981 The social class structure of occupation mobility, in *Amer. J. Soc.*, 87, 578-611.
- Breslow, N.
1981 Odds ratio estimators when the data are sparse, in *Biometrika*, 68, 67-84.
- Brier, S.S.
1980 Analysis of contingency tables under cluster sampling, in *Biometrika*, 67, 591-596.
- Brockett, P.L.
1981 A note on the numerical assignment of scores to ranked categorical data, in *J. Math. Soc.*, 8, 91-101.
- Brown, M.B.
1974 Identification of the sources of significance in two-way contingency tables, in *Appl. Statist.*, 23, 405-413.
- Brown M.B. & Benedetti, J.K.
1977 Sampling behaviour of tests for correlation in two-way contingency tables, in *J. Amer. Statist. Assoc.*, 72, 309-315.
- Burns, P.R.
1984 *Multiple Comparison Methods in MANOVA*, Proceedings 7th SPSS Users and Coordinators Conference.
- Cattell, R.B.
1966 The meaning and strategic use of factor analysis, in Cattell, R.B. (ed.), *Handbook of Multivariate Experimental Psychology*, Chicago, Rand McNally.
- Chen, T.T.
1979 Log-linear models for categorical data with misclassification and double sampling, in *J. Amer. Statist. Assoc.*, 74, 481-488.
- Chen, T.T., Hochberg, Y. & Tenenbein, A.
1984 Analysis of multivariate categorical data with misclassification errors by triple sampling schemes, in *Journal of Statistical Planning Inference*, 9, 177-184.
- Chiacchierini, R.P. & Arnold, J.C.
1977 A two-sample test for independence in 2x2 contingency tables with both margins to misclassification, in *J. Amer. Statist. Assoc.*, 72, 170-174.

- Churchill, G.A., Jr.
1979 *Marketing Research: Methodological Foundations*, Hinsdale, Illinois, Dryden Press.
- Clogg, C.C.
1981 Latent structure models of mobility, in *Amer. J. Soc.*, 86, 836-868.
1982 Some models for the analysis of association in multiway cross-classification having ordered categories, in *J. Amer. Statist. Assoc.*, 77, 803-815.
- Clogg, C.C. & Goodman, L.A.
1984 Latent structure analysis of a set of multidimensional contingency tables, in *J. Amer. Statist. Assoc.*, 79, 762-771.
- Cochran, W.G. & Cox, G.M.
1957 *Experimental Designs*, 2nd ed., New York, John Wiley and Sons.
- Cohen, A., Gastonis, C. & Marden, J.I.
1983 Hypothesis testing for marginal probabilities in a 2x2x2 contingency tables with conditional independence, in *Journal of the American Statistical Association*, 78, 920-929.
- Cohen, J.
1968 Multiple regression as a general data-analytic system, in *Psychological Bulletin* 6, 426-443.
- Coleman, J.S.
1964 *Introduction to Mathematical Sociology*, New York, Free Press.
- Coleman, J.S., Kelly, S.D. & Moore, J.A.
1975 *Recent trends in school integration*, Washington, D.C., Urban Inst.
- Corbetta, P. & Schadee, H.
1984 *L'algoritmo di Stephan-Deming: programmi di calcolo ed applicazioni in sociologia politica*, Relazione presentata al convegno 'Ricerca sociologica, informativa e società italiana', Bari.
- Cortese, C.F., Falk, R.F. & Cohen, J.K.
1976 Further considerations on the methodological analysis of segregation indices, in *American Sociological Review*, 41, 630-637.
- Cox, M.A.A. & Plackett, R.L.
1980 Small samples in contingency tables, in *Biometrika*, 67, 1-14.
- Cox, C.P.
1982 An alternative way of calculating the X² independence or association test statistic for 2xk contingency tables, in *Amer. Statistician*, 36, 133.

- Cressie, R. & Rear, T.R.C.
1984 Multinomial goodness of fit tests, in *J. Roy. Statist. Soc., B*, 46, 440-464.
- Dallal, G.E.
1982 A simple interaction model for two-dimensional contingency tables, in *J. Amer. Statist. Assoc.*, 77, 425-432.
- Darroch, J.N., Lauritzen, S.L. & Speed, T.P.
1980 Markov fields and log-linear interactions models for contingency tables, in *Ann. Stat.*, 8, 522-539.
- Davis, J.A.
1971 *Elementary Survey Analysis*, Englewood, N.J., Prentice Hall.
1974 Hierarchical models for significance tests in multivariate contingency tables: An exegesis of Goodman's recent papers, in Costner, H. (ed.), *Sociological Methodology*, 1973-74, S. Francisco, Jossey Bass.
1975 The log-linear analysis of survey replications, in Land, K. & Spilerman, S. (eds.), *Social Indicators models*, New York, Sage.
1976 Analyzing contingency tables with linear flow graphs: D System, in Heise, D. (ed.), *Sociological Methodology*, 1976, S. Francisco, Jossey Bass.
- De Lillo, A. & Schizzerotto, A.
1982 Diseguaglianze educative ed occupazionali, in Cappello, F., Dei, M. & Rossi, M., *L'immobilità sociale: stratificazione e sistemi scolastici*. Bologna, Il Mulino.
- Duncan, O.D.
1969 Contingencies in constructing causal models, in Borgatta, E.F. (ed.), *Sociological Methodology*, S. Francisco, Jossey Bass.
1975 Partitioning polytomous variables in multiway contingency tables, in *Social Science Research*, 4, 167-182.
1979 How destination depend upon origin in the occupational mobility table, in Schuessler, K.F. (ed.), *Sociological Methodology*, S. Francisco, Jossey Bass.
1979 Testing key hypotheses in panel analysis, in Schuessler, K.F. (ed.), *Sociological Methodology*, S. Francisco, Jossey Bass.
1980 Two faces of panel analysis: parallels with comparative cross-sectionals analysis and time-lagged association, in Leinhardt, S. (ed.), *Sociological Methodology*, S. Francisco, Jossey Bass.
1985 New Light on the 16-fold table, in *American J. Sociology*, 91, 1, 88-128.
- Duncan, O.D. & McRae, J.A., Jr.
1978 Multiway contingency analysis with a scale response or factor, in Schuessler, K.F. (ed.), *Sociological Methodology*, S. Francisco, Jossey Bass.
- Edwards, D. & Krainer, S.
1983 The analysis of contingency tables by graphical methods, in *Biometrics*, 70, 553-565.

- Emerson, I.D. & Moses, L.E.
1985 A note on the Wilcoxon-Mann-Whitney test for 2xk Ordered tables, in *Biometrics*, 41, 303-309.
- Enke, H.
1977 On the analysis of incomplete two-dimensional contingency tables, in *Biom. J.*, 19, 561-573.
1978 On the analysis of spEcial incomplete three dimensional contingency tables, in *Biom. J.*, 20, 229-242.
- Everitt, B.S.
1977 *The analysis of contingency tables*, London, Chapman and Hall.
1977 *Graphical Techniques for Multivariate Data*, New York, North Holland.
1980 *Cluster Analysis*, 2nd ed., London, Heineman Educational Books.
- Evers, M. & Namboodiri, K.
1978 On the design matrix strategy in the analysis of categorical data, in Schuessler, K. (ed.), *Sociological Methodology*, 1979, S. Francisco Jossey Bass.
- Eysenck, M.W.
1977 *Human Memory: Theory, Research and Individual Differences*, New York, Pergamon Press.
- Fellegi, I.P.
1980 Approximate tests of independence and goodness of fit based on stratified multistage, *J. Amer. Statist. Assoc.*, 75, 261-268.
- Fennessey, J.
1968 The general linear model: A new perspective on some familiar topics, in *American Journal of Sociology*, 74, 1-27.
- Fienberg, S.
1977 *The Analysis of Cross-Classified Categorical Data*, Cambridge, MIT Press.
- Fienberg, S.E.
1978 Using log-linear models to analyse cross-classified categorical data, in *Math. Scientist*, 5, 13-30.
1978 A note on fitting and interpreting parameters in models for categorical data, in Schuessler, K.F. (ed.), *Sociological Methodology*, S. Francisco, Jossey Bass.
1979 The use of chi-square statistics for categorical data problems, in *J. Roy. Stat. Soc., B*, 41, 54-64.
- Fienberg, S.E. & Larntz, K.
1976 Log-linear representation for paired and multiple comparison models, in *Biometrika*, 63, 245-254.
- Fingleton, B.
1982 Log-linear models, mostellerizing and forecasting, in *Trans. Inst. Br. Geog.*

- 1982 The family of loyalty-distance models: a log-linear analysis of space-time mobility,
- 1983 Independence, stationarity, categorical spatial data and chi-square test, in *Env. and Plann.*, A, 15, 483-499.
- 1983 Log-linear models with dependent spatial data, in *Env. and Plann.*, A, 15, 801-813.
- Finn, J.D.
1974 *A General Model for Multivariate Analysis*, New York, Holt, Rinehart and Wiston.
- Finney, D.J.
1971 *Probit analysis*. Cambridge, England: Cambridge University Press.
- Fisher, R.A.
1962 Confidence limits for a cross-product ratio, in *Australian Journal Statistics*, 4, 41.
- Formann, A.K.
1981 Log-linear models for polytomous data, in *Biometrical Journal*, 23, 389-404.
- Freund, R.J.
1980 The case of missing cell, *The American Statistician*, 34, 94-98.
- Fuchs, C. & Kenett, R.
1980 A test for detecting outlying cells in the multinomial distribution and two-way contingency tables, in *J. Amer. Statist. Assoc.*, 75, 395-397.
- Gangemi, G.
1980 A critique of Goodman's model for the multivariate analysis of relations between quantitative variables, in *Quality and Quantity*, 2, 343-352.
- Gans, L.P. & Robertson, C.A.
1981 Distributions of Goodman and Kruskal's gamma and Sperman's rho in 2x2 tables for small and moderate sample sizes, in *J. Amer. Statist. Assoc.*, 76, 942-946.
- Gardner, W., Hartmann, D.P. & Mitchell, C.
1982 The effect of serial dependence on the use of chi-squared for analyzing sequential data in dyadic interactions, in *Behavioural Assessment*, 4, 75-82.
- Gart, J.J.
1978 The analysis of ratios and cross-product ratios of Poisson variates with application to incidence rates, in *Comm. Stat. Theor. Meth.*, 11, 1233-1240.
- Gilbert, G.N.
1981 *Modeling society - An Introduction to log-linear analysis for social researchers*, London, Allen & Unwin.

- Gilbert, E.S.
1968 On discrimination using qualitative variables, in *Journal of the American Statistical Association*, 63, 1399-1412.
- Gilula, Z.
1982 A note on the analysis of association in cross-classification having ordered categories, in *Comm. Stat. Theor. Meth.*, 17, 987-1005.
- Goldstein, M. & Dillon, W.R.
1977 A stepwise discrete variable selection procedure, in *Comm. Stat. Theor. Meth.*, A6, 1423-1436.
1978 *Discrete Discriminant Analysis*, New York, Wiley and Sons.
- Goodman, L.A.
1965 On the multivariate analysis of three dichotomous variables, in *American Journal of Sociology*, LXXI, 3, 290-301.
1968 The analysis of cross-classified data: Independence, quasi-independence and interactions in tables with or without missing entries, in *Journal of the American Statistical Association*, 63, 1091-1131.
1970 The multivariate analysis of qualitative data: Interactions among multiple classification, in *J. Amer. Statist. Assoc.*, 65, 226-256.
- Goodman, L.A.
1971 The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications, in *Technometrics*, 13, 33-61.
1972 A general model for the analysis of survey, in *American Journal of Sociology*, 77, 1035-1086.
1972 A modified multiple regression approach to the analysis of dichotomous variables, in *American Sociology Review*, 37, 28-46.
1973 Causal analysis of data from panel studies and other kinds of survey, in *American Journal of Sociology*, 78, 1135-1191.
1976 The relationship between the modified and the more usual multiple regression approach to the analysis of dichotomous variables, in Heise, D. (ed.), *Sociological Methodology*, S. Francisco, Jossey-Bass.
1978 *Analyzing qualitative/categorical data: log-linear models and latent structure analysis*. Reading, Mass., Addison Wesley.
1979 On quasi-independence in triangular contingency tables, in *Biometrics*, 35, 651-656.
1979 Multiplicative models for the analysis of occupational mobility tables and other kinds of cross-classification tables, in *American J. Sociology*, 84, 804-819.
1979 A brief guide to the causal analysis of data from survey, in *The American Journal of Sociology*, 84, 1078-1095.
1979 Multiplicative models for square contingency tables with ordered categories, in *Biometrika*, 66, 413-418.
1981 Association models and the bivariate normal for contingency tables with ordered categories, in *Biometrika*, 68, 347-355.

- 1981 Association models for canonical correlation in the analysis of cross-classification having ordered categories, in *J. Amer. Statist. Assoc.*, 76, 320-334.
- 1981 Criteria for determining whether certain categories in a cross-classification table should be combined, with special reference to occupational categories in an occupational mobility table, in *Amer. J. Soc.*, 87, 612-650.
- 1983 The analysis of dependence in cross-classification having ordered categories, using log-linear models for frequencies and log-linear models for odds, in *Biometrics*, 39, 149-160.
- 1984 *The analysis of cross-classified data having ordered categories*, Cambridge, Harvard University Press.
- 1985 The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries, in *Annals of Statistics*, 13, 10-69.
- Goodman, L.A. & Kruskal, W.H.
1954 Measures of association for cross-classifications, in *Journal of the American Statistical Association*, 49, 732-764.
- Gower, J.C.
1977 The analysis of asymmetry and orthogonality, in *Recent Developments in Statistics*, (Barra, J.R. et al., eds.), Amsterdam, North Holland.
- Gray, L.N. & Williams, J.S.
1981 Goodman and Kruskal's tau b: multiple and partial analogs, in *Soc. Meth. & Res.*, 10-50-62.
- Green, B.F.
1979 The two kinds of linear discriminant functions and their relationship, in *Journal of Educational Statistics*, 4:3, 247-263.
- Green, P.E.
1978 *Analyzing Multivariate Data*, Hinsdale, Illinois, Dryden Press.
- Greenhouse, S.W. & Geisser, S.
1959 On methods in analysis of profile data, in *Psychometrika*, 24, 95-112.
- Grizzle, J.E., Starmer, C.F. & Koch, G.G.
1969 Analysis of categorical data by linear models, in *Biometrics*, 25, 489-504.
- Gross, W.F.
1984 A note on "Chi-square tests with survey data", in *J. Roy. Statist. Soc.*, B, 46, 270-272.
- Gross, S.T.
1981 On the asymptotic power and efficiency of tests of independence in contingency tables with ordered classifications, in *J. Amer. Statist. Assoc.*, 76, 935-941.

- Grove, D.M.
1980 A test on independence against a class of ordered alternatives in a 2xc contingency table, in *J. Amer. Statist. Assoc.*, 75, 454-459.
- 1984 Positive association in a two-way contingency table: likelihood-ratio tests,
- Haber, M.
1980 A comparison of some continuity corrections for the chi-square test on 2x2 tables, in *J. Amer. Statist. Assoc.*, 75, 510-515.
- 1982 Testing for independence in intraclass contingency tables, in *Biometrics*, 38, 93-104.
- 1983 Multidimensional intraclass contingency tables, in *Comm. Statist. Theor. Meth.*, 12, 765-776.
- 1984 A comparison of tests for the hypothesis of no three-factor interaction in 2x2x2 contingency tables, in *J. Statist. Comp. Sim.*, 20, 205-215.
- 1984 Fitting a general log-linear model: AS207, in *Appl. Statist.*, 33, 358-362.
- Haberman, S.J.
1973 The analysis of residuals in cross-classified tables, in *Biometrics*, 29, 205-220.
- 1974 *The analysis of frequency data*, Chicago, University of Chicago Press.
- 1974b Log-linear models for frequency tables with ordered classifications, in *Biometrics*, 39, 589-600.
- 1978 *Analysis of qualitative data. Vol. 1: Introductory topics*, New York, Academic Press.
- 1979 *Analysis of qualitative data. Vol. 2: New developments*, New York, Academic Press.
- 1981 Test for independence in two-way contingency tables based on canonical correlation and on linear-by-interactions, in *Ann. Stat.*, 9, 1178-1186.
- Hand, D.J.
1981 *Discrimination and Classification*, New York, Wiley and Sons.
- Harman, H.H.
1967 *Modern factor analysis*, 2nd ed., Chicago, University of Chicago Press.
- Hauck, W.W., Anderson, S. & Leahy, F.J.
1982 Finite sample properties of some old and some new estimators of a common odds-ratio from multiple 2x2 tables, *J. Amer. Statist. Assoc.*, 77, 145-152.
- Hauck, W.W.
1979 The large-sample variance of the Mantel-Haenszel estimator of a common odds-ratio, in *Biometrics*, 35, 817-819.
- Hauser, R.M.
1978 A structural model of the mobility table, in *Social Forces*, 56, 919-953.
- 1979 Some exploratory methods for modeling mobility tables and other cross-classified data, in Schuessler, K.F. (ed.), *Sociological Methodology*, 1980, S. Francisco, Jossey Bass.

- Havranek, T.
1984 A procedure for model search in multidimensional contingency tables, in *Biometrics*, 40, 95-100.
- Heilbron, D.C.
1981 The analysis of ratios of odds ratios in stratified contingency tables, in *Biometrics*, 37, 55-56.
- Heise, D.R.
1975 *Causal Analysis*, New York, Wiley and Sons.
- Henry, N.W.
1981 Jackknifing measures of association, in *Soc. Meth. Res.*, 10, 233-240.
- Holt, D.
1979 Log-linear models for contingency tables analysis: an interpretation of parameters, in *Soc. Meth. Res.*, 7, 330-336.
- Hope, K.
1981 Vertical mobility in Britain: a structured analysis, in *Sociology*, 15, 19-55.
1982 Vertical and nonvertical class mobility in three countries, in *Amer. Soc. Rev.*, 47, 99-113.
- Hosmer, D.W. & Hartz, S.C.
1981 Methods for analyzing odds ratios in a 2xc contingency table, in *Biometrical Journal*, 23, 741-748.
- Horton, R.L.
1978 *The general linear model*, New York, McGraw-Hill.
- Hubert, L.J. & Golledge, R.G.
1982 Measuring association between spatially defined variables: Tjostheim's index and some extensions, in *Geographical Analysis*, 14, 273-278.
- Huynh, H. & Feldt, L.S.
1976 Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs, in *Journal of Educational Statistics*, 1, 69-82.
- Imrey, P.B., Koch, G.G. & Stokes, M.E.
1981 Categorical data analysis: some reflections on the log-linear model and logistic regression. Part I: Historical and methodological overview, in *Int. Stat. Rev.*, 49, 265-283.
1982 Categorical data analysis: some reflections on the log-linear model and logistic regression. Part II: Data analysis, in *Int. Stat. Rev.*, 50, 35-63.

- Janson, S. & Vegelius, J.
1982 Correlation coefficients for more than one scale type, in *Mult. Behav. Res.*, 17, 271-284.
- Johnston, J.
1972 *Econometric Methods*, 2nd ed., New York, McGraw-Hill.
- Jonassen, C.T. & Peres, S.H.
1960 *Interrelationship of Dimensions of Community Systems*, Columbus, Ohio University Press.
- Kaiser, H.F.
1974 An index of factorial simplicity, in *Psychometrika*, 39, 31-36.
- Kannemann, K.
1982 The exact evaluation of 2-way cross-classification: an algorithmic solution, in *Biometrical Journal*, 24, 157-169.
- Kass, G.V.
1980 An exploratory technique for investigating large quantities of categorical data, in *Appl. Stat.*, 29, 119-127.
- Kempthorne, O.
1979 In dispraise of the exact test: a rejoinder, in *J. Statist. Plann. Inf.*, 3, 199-213.
- Khamis, H.J.
1983 Log-linear model analysis of the semi-symmetric intraclass contingency table, in *Communications in Statistics: Theory and Methods*, 12, 2723-2752.
- Kiiveri, H. & Speed, T.P.
1982 Structural analysis of multivariate data: a review, in Leinhardt, S. (ed.), *Sociological Methodology*, S. Francisco, Jossey Bass.
- Kim, J.O. & Mueller, C.W.
1978 *Introduction to factor analysis*, Beverly Hills, Sage Press.
- Knoke, D.
1974 Religious involvement and political behavior: A Log-linear analysis of white Americans 1952-1968, in *Sociological Quarterly*, 3, 416-434.
1975 A comparison of log-linear and regression models for systems of dichotomous variables, in *Sociological Methods and Research*, 3, 416-434.
- Knoke, D. & Burke, P.J.
1980 *Log-linear models*, Beverly Hills, Sage Publications.

- Koehler, K.J. & Larntz, K.
1980 An empirical investigation of goodness-of-fit statistics for sparse multinomials, in *J. Amer. Statist. Assoc.*, 75, 336-344.
- Korn, E.I.
1981 Hierarchical log-linear models not preserved by classification errors, in *J. Amer. Statist. Assoc.*, 76, 110-113.
1982 The asymptomatic efficiency of tests using misclassified data in contingency tables, in *Biometrics*, 38, 445-450.
- Kotze, T.J.V. & Hawkins, D.M.
1984 The identifications of outliers in two-way contingency tables using 2x2 subtables, in *Appl. Stat.*, 33, 215-223.
- Kousser, J.M., Cox G.W. & Galenson, D.W.
1982 Log-linear analysis of contingency tables: An introduction for historians with an application to Thernstrom on the "Floating Proletariat", in *Historical Methods*, 15, 152-169.
- Krewski, D., Brennan, J. & Bickis, M.
1984 The power of the Fisher permutation test in 2xk tables, in *Comm. Stat. - Sim. Comp.*, 13, 433-448.
- Kritzer, H.M.
1979 Ordinal measures of association, in *Journal of the American Statistical Association*, 53, 814-861.
- Kshirsagar, A.M. & Arseven, E.
1975 A note on the equivalency of two discrimination procedures, in *The American Statistician*, 29, 38-39.
- Kviz, F.J.
1981 Interpreting proportional reduction in error measures as percentage of variation explained, in *Soc. Quarterly*, 22, 413-420.
- Lachenbruch, P.A.
1975 *Discriminant Analysis*, New York, Hafner Press.
- Laird, N.M.
1978 Empirical Bayes methods for two-way contingency tables, in *Biometrika*, 65, 581-590.
- Land, K.C.
1969 Principles of path analysis, in E.F. Borgotta (ed.), *Sociological Methodology*, S. Francisco Jossey Bass.

- Landis, J.R. & Koch, G.G.
1977 A one way components of variance model for categorical data, in *Biometrics*, 33, 671-680.
- Landis, J.R., Heyman, E.R. & Koch, G.G.
1978 Average partial association in three-way contingency tables: a review and discussion of alternative tests, in *Int. Statist. Rev.*, 46, 237-254.
- Larntz, K.
1978 Small sample comparison of exact levels for chi-square goodness of-fit statistics, in *J. Amer. Statist. Assoc.*, 73, 253-263.
- Lawal, H.B.
1984 Comparison of the X², Y², Freeman-Tucheyt and William's improved G² test statistics in small samples in one-way multinomials, in *Biometrika*, 71, 415-418.
- Lawal, H.B. & Upton, G.J.G.
1980 An approximation to the distribution of the X² goodness-of-fit statistic for use with small expectations, in *Biometrika*, 67, 447-454.
1984 On the use of X² as a test of independence in contingency tables with small cell expectations, in *Australian Journal of Statistics*, 26, 75-85.
- Lee, S.K.
1977 On the asymptotic variances of u terms in log-linear models of multidimensional contingency tables, in *J. Amer. Statist. Assoc.*, 72, 412-419.
- Lefebvre, J.
1976 *Introduction aux analyses statistiques multidimensionnelles*, Paris, Masson.
- Leonard, T.
1975 Bayesian estimation methods for two-way contingency tables, in *J. Roy. Statist. Soc.*, B, 37, 23-37.
- Lever, H.
1978 Goodman's methods of analysis, in *Humanitas*, 4, 361-371.
- Lewis, T., Saunders, I.W. & Westcott, M.
1984 The moments of the Pearson chi-squared statistic and the minimum expected value in two-way tables, in *Biometrika*, 71, 515-522.
- Lindsey, Y.K.
1973 *Inference from sociological survey data. A unified approach*, New York, Elsevier.
- Lingoes, J.C.
1968 The multivariate analysis of qualitative data, in *Multivariate Behavioral Research*, January, 81-93.

- Long, J.S.
1984 Estimable functions in log-linear models, in *Soc. Meth. Res.*, 12, 399-432.
- Maghsoodloo, S.
1981 The continuity correction cor chi-square for 2x2 tables, in *J. Statist. Comp. Simul.*, 14, 47-79.
- Magidson, J., Swan, J.H. & Berk, R.A.
1981 Estimating non-hierarchical and nested log-linear models, in *Soc. Meth. Res.*, 10, 3-49.
- Mantel, N. & Byar, D.P.
1978 Marginal homogeneity, symmetry and independence, in *Comm. Statist. Theor. Meth.*, A7, 953-976.
- Marascuilo, L.A. & Dagenais, F.
1982 Planned and post hoc comparison for tests of homogeneity where the dependent variable is categorical and ordered, in *Educational and Psychological Measurement*, 42, 777-781.
- McCaa, R.
1982 Modelling social interaction: marital miscegenation in colonial Spanish America, In *Historical Methods*, 15, 45-63.
- McCullagh, P.
1978 A class of parametric models for the analysis of square contingency tables with ordered categories, in *Biometrika*, 65, 413-418.
1980 Regression models for ordinal data, in *J. Roy. Statist. Assoc.*, B, 42, 109-142.
1982 Some applications of quasisymmetry, in *Biometrika*, 69, 303-308.
1984 On the elimination of nuisance parameters in the proportional odds model, in *J. Roy. Statist. Assoc.*, B, 46, 250-256.
1985 On the asymptotic distribution of Pearson's statistic in liner exponential-family models, in *International Statistical Review*, 53, 61-67.
- McDonald-Schlichting, U.
1979 Note on simply calculating chi-square for rxc contingency tables, in *Biom. J.*, 21.
- Meeks, S.L. & D'Agostino, R.B;
1983 A model for comparison with ordered categorical data, in *Comm. Statist. Theor. Meth.*, 12, 895-906.
- Miller, R.G.
1981 *Simultaneous statistical inference*, 2nd ed., New York, Springer Verlag.
- Milligan, G.W.
1980 An examination of the effect of six types of error perturbation on fifteen clustering algorithms, in *Psychometrika*, 45, 325-342.

- Milligan, G.W. & Isaac, P.D.
1980 The validation of four ultrametric clustering algorithms, in *Pattern Recognition*, 12, 41-50.
- Milliken, G.W. & Johnson, D.E.
1984 *Analysis of Messy Data*, Belmont, California, Lifetime Learning Publications.
- Mood, A., Mac Farlane & Graybill, F.A.
1963 *Introduction to the theory of statistics*, New York, McGraw-Hill.
- Moore, D.H.
1973 Evaluation of five discrimination procedures for binary variables in *Journal of the American Statistical Association*, 68, 399.
- Morgan, J.N. & Lonquist, J.A.
1963 Problems in the analysis of survey data and a proposal, in *J. Amer. Statist. Assoc.*, 58, 302, 415-434.
- Morgan, B.J.T. & Titterington, D.M.
1977 A comparison of iterative methods for obtaining maximum likelihood estimates in contingency tables with missing diagonal, in *Biometrika*, 64, 26-269.
- Morrison, D.F.
1967 *Multivariate Statistical Methods*, New York, McGraw-Hill.
- Mosteller, F.
1968 Association and estimation in contingency tables, in *J. Amer. Statist. Assoc.*, 63, 1-28.
- Nelder, J.A.
1974 Log-linear models for contingency tables: A generalization of classical least squares, in *Applied Statistics*, 3, 323-329.
- Norusis, M.J.
1983 *SPSSx Introductory Statistics Guide*, New York, McGraw-Hill.
1985 *SPSSx Advanced Statistics Guide*, New York, McGraw-Hill.
- Oler, J.
1985 Non centrality parameters in chi-squared goodness-of-fit analysis with an application to log-linear procedures, in *J. Amer. Statist. Assoc.*, 80, 181-189.
- Olsen, C.L.
1976 On choosing a test statistic in multivariate analysis of variance, in *Psychological Bulletin*, 83, 579-586.
- O'Muircheartaigh, C.A. & Payne, C. (eds.)
1977 *The analysis of survey data*, London, Wiley.

- O'Neill, M.E.
1980 The distribution of higher-order interactions in contingency tables, in *J. Roy. Statist. Assoc.*, B, 42, 357-365.
- Pagano, M. & Halvorsen, K.T.
1981 An algorithm for finding the exact significance levels of rxc tables, in *J. Amer. Statist. Assoc.*, 76, 931-934.
- Patefield, W.M.
1982 Exact tests for trends in ordered contingency tables, in *Appl. Statist.*, 31, 32-43.
- Payne, C.
1977 Log-linear models for contingency tables, in Muirheartaigh, C.A. & Payne, C. (eds.), *The analysis of survey data*, Vol. 2: Model Fitting, New York, Wiley.
- Pettitt, A.N.
1984 Tied, grouped continuous and ordered categorical data: a comparison of two models, in *Biometrika*, 71- 35-42.
- Plackett, R.L.
1974 *The analysis of categorical data*, London, Giffin.
- Plackett, R.L. & Paul, S.R.
1978 Dirichlet models for square contingency tables, in *Comm. Stat. - Theor. Meth.*, A7, 939-952.
- Rao, J.K. & Scott, A.J.
1981 The analysis of categorical data from complex sample survey: chi-square tests of goodness-of-fit and independence in two-way tables, in *J. Amer. Statist. Assoc.*, 76, 221-230.
- Reynolds, H.T.
1977 *The analysis of nominal data*, London, Sage.
1977 *The analysis of cross-classification*, New York, Free Press.
1978 Some comments on the causal analysis of surveys with log-linear models, in *Amer. J. Soc.*, 83, 127-143.
- Rhoades, H.M. & Overall, J.E.
1982 A sample size correction for Pearson chi-square in 2x2 contingency tables, in *Psychol. Bull.*, 91, 418-423.
- Romesburg, H.C.
1984 *Cluster analysis for researchers*, Belmont, California, Lifetime Learning Publications.

- Roy Samarendra, N. & Kastenbaum, M.A.
1956 Hypothesis of no-interaction in a multi-way contingency table, in *The Annals of Mathematical Statistics*, 27, 749-757.
- Santner, T.J. & Snell, M.K.
1980 Small sample confidence intervals for p1-p2 in 2x2 contingency tables, in *J. Amer. Statist. Assoc.*, 75, 386-394.
- Searle, S.R.
1971 *Linear models*, New York, Wiley and Sons.
- Semenya, K.A., Koch, G.G. & Stokes, M.E.
1983 Linear models for some rank function analyses of ordinal categorical data, in *Comm. Statist. Theor. Meth.*, 12, 1277-1298.
- Shaffer, J.P.
1977 Reorganization of variables in analysis of variance and multidimensional contingency tables, in *Psychol. Bull.*, 84, 220-228.
- Simon, G.
1974 Alternative analysis for the singly ordered contingency table, in *J. Amer. Statist. Assoc.*, 69, 971-976.
1978 Efficacies of measures of association for ordinal contingency tables, in *J. Amer. Statist. Assoc.*, 73, 545-551.
- Simpson, E.H.
1951 The interpretation of interaction in contingency tables, in *J. Roy. Statist. Soc.*, B, 13, 238-241.
- Sneath, P.H.A. & Sokal, R.R.
1973 *Numerical Taxonomy*, S. Francisco, W.H. Freeman and Co.
- Somers, R.H.
1980 Simple approximation to null sampling variances, in *Soc. Meth. and Res.*, 9, 115-126.
- SPSS Inc.
1985 *SPSS Statistical Algorithms*, Chicago: SPSS Inc.
1986 *SPSSx User's Guide*, 2nd ed., New York, McGraw-Hill.
1986 *SPSS/PC for the IBM PC/XT/AT*, M.J. Norusis, Chicago: SPSS Inc.
- Stephan, F.F. & Deming, W.E.
1940 On a least squares adjustment of sampled frequency table when the expected marginal totals are known, in *The Annals of Mathematical Statistics*, n. 11.

- Stoetzel, J.
1960 A factor analysis of liquor preference of french consumers, in *Journal of Advertising Research*, 1: 1, 7-11.
- Stinchombe, A.L.
1983 Linearity in log-linear analysis, in Leinhardt, S. (ed.), *Sociological Methodology 1983/1984*, S. Francisco, Jossey Bass.
- Subrahmaniam, K.
1979 Test of significance for the odds ratio in 2x2 table based on the ratio of two independent F-ratios, in *Comm. Stat. Simula. Computa.*, B, 8, 245-255.
- Swafford, M.
1980 Three parametric techniques for contingency table analysis: a nontechnical commentary, in *American Sociological Review*, 45, 664-690.
- Tatsuoka, M.M.
1971 *Multivariate Analysis*, New York, Wiley and Sons.
- Tavare, S.
1983 Serial dependence in contingency tables, in *J. Roy. Statist. Assoc.*, B, 45, 100-106.
- Tavare, S. & Altham, P.M.E.
1983 Serial dependence of observations leading to contingency tables, and corrections to chi-square statistics, in *Biometrika*, 70, 139-144.
- Taylor, D.G.
1983 Analyzing qualitative data, in Rossi, P., Smith, J. & Anderson, A.B., *Handbook of survey research*, New York, McMillan.
- Theil, H.
1970 On the estimation of relationship involving qualitative variables, in *American Journal of Sociology*, 76, 103-154.
- Thomas, D.G. & Gart, J.J.
1977 A table of exact confidence limits for differences and ratios of two proportions and their odds ratios, in *J. Amer. Statist. Assoc.*, 72, 73-76.
- Timm, N.H.
1975 *Multivariate analysis with applications in education and psychology*, Monterey, Brooks/Cole.
- Tjostheim, D.
1978 A measure of association for spatial variables, in *Biometrika*, 65, 109-114.

- Tucker, L.R.
1971 Relations of factor score estimates to their use, in *Psychometrika*, 36, 427-436.
- Tucker, R.F., Koopman, R.F. & Linn, R.L.
1969 Evaluation of factor analytic research procedures by means of simulated correlation matrices, in *Psychometrika*, 34, 421-459.
- Upton, G.
1978 *The analysis of cross-tabulated data*, New York, Wiley.
1981 Log-linear models, screening and regional industrial surveys, in *Reg. Studies.*
1982 A comparison of alternative tests for the 2x2 comparative trial, in *J. Roy. Statist. Assoc.*, A, 145, 86-105.
1982 Categorical data, in Kotz, S. & Johnson, N.L. (eds.), *Encyclopaedia of Statistical Sciences*, Amsterdam, North Holland.
1984 Modelling cross-tabulated regional data, in Nijkamp, P.J. (ed.), *Measuring the Unmeasurable*, Amsterdam, North Holland.
- Upton, G. & Sarlvik, B.H.
1981 A loyalty-distance model for voting change, in *J. Roy. Statist. Assoc.*, A, 144, 247-259.
- Van Vliet, P.K.J. & Gupta, J.M.
1973 THAM v.sodium bicarbonate in idiopathic respiratory distress syndrome, in *Archives of disease in childhood*, 48, 249-255.
- Wahl, P.W. & Kronmal, R.A.
1977 Discriminant functions when covariances are unequal and sample size are moderate, in *Biometrics*, 33, 479-484.
- Wermuth, N. & Lauritzen, S.L.
1983 Graphical and recursive models for contingency tables, in *Biometrika*, 70, 537-552.
- Weima, J.
1979 Tests of independence and coefficients for RxC contingency tables, in *Quality & Quantity*, 13, 255-265.
- White, A.A., Landis, J.R. & Cooper, M.M.
1982 A note on the equivalence of several marginal homogeneity test criteria for categorical data, in *Int. Statist. Rev.*, 50, 27-34.
- Whittemore, A.S.
1978 Collapsibility of multidimensional contingency tables, in *J. Roy. Statist. Soc.*, B, 40, 328-340.
- Winer, B.J.
1971 *Statistical principles in experimental designs*, New York, McGraw-Hill.

Witkin, H.A. et Al.

1954 *Personality through perception*, New York, Harper.

Woolfson, R.F. & Brier, S.S.

1981 Equivalences of certain chi-squared statistics, in *Amer. Statist.*, 35, 250-253.

Wrigley, N.

1980 Paired comparison experiments and logit models: a review and illustration of recent developments, in *Env. and Plann.*, A, 12, 21-40.

1980 Categorical data, repeated measurements research designs and regional industrial surveys, in *Regional Studies*, 14, 455-471.

Zahn, D.A. & Fein, S.B.

1979 Large contingency tables with large cell frequencies: a model search algorithm and alternative measures of fit, in *Psycho. Bull.*, 86, 1189-1200.

Zidek, J.

1984 Maximal Simpson-disaggregations of 2x2 tables, in *Biometrika*, 71, 187-190.

Appendice

APPENDICE N° 1.*

La ricerca longitudinale sul proseguimento degli studi dopo la scuola dell'obbligo in Provincia di Trento (1973/74 - 1980/81)

La ricerca si è posta l'obiettivo di studiare in modo diretto e dettagliato il comportamento scolastico di un'intera leva di allievi residenti nella provincia di Trento, rilevato ad un momento iniziale di frequenza alla terza media e seguito poi, longitudinalmente, lungo l'intero arco di studi superiore. Il periodo in cui si è svolta la rilevazione include pertanto l'arco di tempo che va dall'anno scolastico 1973/74 - anno della rilevazione iniziale - all'anno scolastico 1980/81, anno in cui tutti gli allievi hanno terminato gli studi regolarmente o in ritardo di uno o due anni¹.

I metodi di rilevazione e di stima

L'approccio seguito per stimare il comportamento scolastico effettivo della leva di allievi è stato duplice. Da un lato si è somministrato un questionario nominativo annuale a tutti gli studenti frequentanti i vari livelli del sistema scolastico considerato, vale a dire la classe III media e i vari anni della scuola secondaria post-obbligo della provincia di Trento, comprensiva dell'istruzione professionale. Tale questionario, riportato in allegato, era assai semplice e includeva, accanto ai dati sulla situazione scolastica attuale e passata dello studente, alcune informazioni di carattere anagrafico, sociale e territoriale (sesso, età, comune di residenza, condizioni sociali, culturali e professionali dei genitori), che hanno consentito l'analisi dei comportamenti, degli status e dei percorsi scolastici separatamente per maschi e femmine, per i vari livelli sociali, per allievi

* Si ringraziano i Dott. Francesco Merz e Luciano Sommacal per il contributo fornito nella compilazione delle tabelle pubblicate alla fine di questa appendice.

¹ La prima fase di questa ricerca, relativa al passaggio dalla terza media alla prima classe del sistema scolastico post-obbligo trentino, è stata pubblicata nel 1976: G. Chiari, A. de Lillo, *Il proseguimento degli studi dopo la scuola dell'obbligo in Provincia di Trento*, Padova, CLEUP, in "Problemi e metodi di programmazione scolastica nella prospettiva del distretto", pp. 215-232.

residenti nei centri urbani e non urbani.

Un secondo approccio, parallelo a quello appena descritto, ha raccolto i dati relativi allo status scolastico dei singoli allievi (regolarità, ripetenze, abbandoni...) nelle segreterie delle varie sedi scolastiche, attraverso un modello di rilevazione semplificato (anch'esso riportato in allegato) rivolto essenzialmente a ricostruire il percorso scolastico regolare e irregolare (con ritardi, trasferimenti, rientri, abbandoni).

Dalle informazioni ottenute con i due metodi sono state costruite le tabelle qui riportate.

Una stima globale degli aggregati scolastici classici: regolarità e ripetenze, promozioni e bocciature, ritardi e abbandoni.

Dato il carattere longitudinale delle rilevazioni e l'approccio per generazioni adottato, è stato possibile stimare anno per anno i vari tassi di passaggio, di regolarità, di ripetenze, di rientri e di abbandoni, di promozione e di bocciatura, di proseguimento e di ripetenza, secondo la triplice ottica della provenienza, esito e destinazione.

Le tabelle n. 3 e 4 riportano i dati (valori assoluti e percentuali) relativi ai diversi comportamenti e status scolastici riscontrati nei cinque livelli del sistema scolastico post-obbligo provinciale.

Gli iscritti

Il numero di studenti iscritti tende a calare linearmente con l'aumentare del livello scolastico considerato; si passa da 6.927 licenziati in III media nell'anno scolastico 1973/74 a 5.258 studenti iscritti in prima nell'anno scolastico 1974/75 ed a soli 1.930 iscritti ad una classe quinta superiore nel 1978/79, il che equivale a dire che su 100 alunni licenziati in III media 75.8 risultano iscritti, un anno dopo, in una classe prima superiore o professionale e soltanto 27.8 arrivano a frequentare una classe quinta della scuola secondaria superiore.

Guardando all'interno del sistema scolastico post-obbligo, la percentuale di alunni che permane nel sistema scolastico trentino dopo 5

anni è la seguente:

	Totale	Maschi	Femmine
Iscritti in I	100	100	100
Iscritti in V cinque anni dopo	36.7	38.9	34.1

Regolarità e permanenza

Le tabelle 3 bis e 3 ter riportano i tassi netti di proseguimento regolare ad una classe alla successiva per tutto l'arco scolastico post-obbligo: la percentuale di alunni che proseguono regolarmente in seconda è 84.6 (81.4% per i maschi; 88.2 per le femmine); in terza è 74.5; in quarta scende a 57.5, in quinta risale a 79.3 (86.9% per i maschi; 70.9% per le femmine). La percentuale di alunni permanenti nel sistema scolastico, sia in regola che come ripetenti è lievemente superiore, comprendendo la percentuale di alunni che, bocciati, ripetono la classe (cfr. tab. 3 bis).

Qualora si voglia costruire, sulla base dei dati disponibili, una vera e propria tavola di regolarità o di permanenza scolastica si ottengono i dati riportati nelle tabelle 3 bis e 3 ter già citate e qui brevemente sintetizzati.

	I	II	III	IV	V
Tavola di regolarità	100	84.6	63.0	36.2	28.7
Tavola di permanenza	100	94.4	75.0	46.7	39.0

In altri termini, la percentuale (o la probabilità) di alunni che, iscritti nel 1974/75 in una classe prima del sistema scolastico post-obbligo della provincia di Trento, ha di continuare regolarmente e di frequentare dopo cinque anni una classe quinta è 28.7%. Tale percentuale è pressochè identica per maschi e femmine; viceversa, la percentuale (o la probabilità) di permanenza nel sistema scolastico post-obbligo risulta superiore per i maschi (43.6) che per le femmine (32.1).

Quando si considerino le differenze per sesso, la caduta numerica degli

iscritti al sistema scolastico post-obbligo con il passare degli anni appare maggiore per le ragazze (da 100 a 34.1) che per gli studenti maschi (da 100 a 38.9).

Regolari e ripetenti

I tassi globali di regolarità sono riportati dalla Tabella 4, quelli distinti per maschi e femmine dalla Tabella 6. Complessivamente la regolarità risulta abbastanza elevata (84.3% nelle classi prime, 95% nelle terze, 92.1% nelle classi quinte) con una leggera migliore situazione iniziale delle femmine rispetto ai maschi (86.7% contro 82.1 nelle prime).

I tassi di ripetenza nelle classi prime (sempre dalle tabelle 4 e 6) ammontano globalmente al 9.8% (12.4% i maschi; 7.0% le femmine); lievemente superiore alle medie nazionali, per gli anni successivi, viceversa, i tassi di ripetenza risultano inferiori a quelli medi nazionali. Come risulta dal seguente prospetto:

Tassi di ripetenza nella scuola secondaria superiore (anno scolastico 1974/75):

	I	II	III	IV	V
Provincia di Trento	9.8	5.0	4.8	4.0	2.9
Italia ¹	8.2	7.7	7.0	5.0	4.3

In generale i tassi di ripetenza dei maschi risultano quasi doppi di quelli delle femmine, in tutti gli anni di corso.

I rientri

Il modello di ricerca adottato ha consentito la stima degli alunni rientrati

¹ Cfr. Censis, *La situazione educativa del paese*, Quindicinale di note e commenti, 1980, XVI, p. 73.

nel sistema scolastico dopo uno o più anni di temporaneo abbandono o dopo uno o più 'anni intermedi'.

Il fenomeno, come era prevedibile, riguarda in misura superiore le femmine dei maschi, quasi esclusivamente nelle classi prime dei vari cicli di studi. I rientri, infatti, che complessivamente ammontano al 2.4% degli iscritti in prima; riguardano solo l'1.6% degli studenti maschi e il 3.3% delle femmine.

Proseguimento ed abbandono

Una volta entrati nel sistema scolastico post-obbligo, gli studenti della provincia di Trento tendono a portare a termine gli studi. I tassi di abbandono sono assai ridotti, specialmente dopo il superamento del primo anno, alla fine del quale si verifica la massima selezione. Complessivamente, su 100 alunni iscritti in prima, 9.4 abbandonano alla fine dell'anno (11.6% i maschi; 7.0% le femmine).

Vi è da dire che i tassi nazionali, calcolati dal CENSIS per l'anno scolastico 1974/75 corrispondente, risultano assai superiori (16.2%) per quanto riguarda l'abbandono dopo la classe prima; pressochè simili, per quanto riguarda il passaggio dalla seconda alla terza (7.8% il tasso nazionale; 6.5% il tasso in provincia di Trento).

Per i dati completi, relativi ai cinque anni di corso, rimandiamo ancora alle tabelle 4 e 6 ed alla figura n. 1.

Promozione e bocciatura

La selezione scolastica formale, che trova nei tassi di bocciatura il suo classico indicatore, risulta alquanto ridotta e tende a diminuire ulteriormente con l'aumentare dei livelli scolastici. La percentuale complessiva di bocciati è pari al 12.7% alla fine del primo anno della scuola post-obbligo. In generale i maschi risultano più bocciati delle femmine: le percentuali di alunni maschi bocciati, corrispondenti ai vari anni di corso, risultano quasi doppie di quelle delle studentesse, come

risulta dalla seguente tabella:

	I	II	III	IV	V
Maschi	16.3	11.9	11.3	9.6	6.7
Femmine	8.6	7.2	6.9	4.8	3.4
Totale	12.7	9.7	9.4	7.3	5.3

Oltre ai dati appena analizzati, la Tabella 7 riporta le percentuali di alunni bocciati che ripetono la classe e di coloro che, viceversa, abbandonano la scuola dopo la bocciatura.

Mentre la percentuale di alunni che ripetono rimane pressochè costante nei vari anni di corso, stabilizzata attorno al 5%, gli abbandoni tendono a diminuire notevolmente con il progredire del corso di studi, tanto è vero che, del 12.7% di alunni bocciati in prima più della metà (58.4%) abbandona la scuola. In quarta e quinta, viceversa, questa percentuale scende pressochè a zero. Anche qui le differenze fra maschi e femmine risultano significative: dopo la bocciatura i maschi tendono ad abbandonare di più all'inizio del corso di studi, di meno alla fine di esso (cfr. Tab. 7 e la fig. n. 3).

Diplomati e qualificati

Le percentuali di alunni che, iscritti ai vari anni della scuola secondaria post-obbligo della provincia di Trento, abbandonano gli studi col titolo finale sono riportate in Tabella 8. Complessivamente il 13.9% consegue un diploma di qualifica dopo due anni di corso. Tra questi, le femmine prevalgono nettamente sui maschi (21.4% contro il 7.0%); dopo tre anni di corso si diplomano il 31.9% di allievi iscritti al terzo anno; dopo quattro anni conseguono un diploma di maturità il 12.9% (2.9% i maschi, 23.9% le femmine, il che è assai plausibile, trattandosi esclusivamente di diplomi magistrali); dopo cinque anni la percentuale di alunni iscritti che conseguono il diploma di maturità è del 94.5% (93.0% per i maschi; 96.5% per le femmine). Complessivamente, il tasso di realizzazione del

sistema post-obbligo in provincia di Trento (definito qui come rapporto fra alunni diplomati e alunni iscritti in prima, cinque anni prima) è pari al 75.8%.

I percorsi irregolari

L'approccio longitudinale seguito ha permesso di stimare alcuni comportamenti scolastici atipici, in particolare i trasferimenti ed i cambiamenti di sede scolastica, tipo di studi e provincia, che normalmente sfuggono alle statistiche ordinarie.

Complessivamente, si è potuto stimare nell'ordine del 9.3% l'ammontare di percorsi irregolari corrispondenti al passaggio dal primo al secondo anno di studi post-obbligo. Tale valore è il massimo riscontrato nell'intero arco quinquennale considerato. Alla fine della terza e della quarta si verificano cambiamenti e trasferimenti nell'ordine del 5%.

In particolare fra gli iscritti delle classi prime, l'incidenza di tali percorsi irregolari risulta così distribuita:

Hanno cambiato:	V.A.		
	Regolari	Ripetenti	Totale
1. Istituto	141	223	364
2. Tipo di scuola	128	156	284
3. Provincia	136	16	152
	%		
	Regolari	Ripetenti	Totale
1. Istituto	3.2	43.0	6.9
2. Tipo di scuola	2.9	30.1	5.4
3. Provincia	3.1	3.1	2.9

Come si vede, se per gli alunni regolari il cambiamento rappresenta l'eccezione, un evento raro che riguarda il 3% di soggetti, per i ripetenti il cambiare sede scolastica e indirizzo di studi appare un comportamento

assai generalizzato. Degli alunni ripetenti che frequentano una qualunque classe prima in provincia di Trento, più del 30% risulta aver cambiato tipo di studi e più del 43% la sede scolastica. E' un dato, questo, netto e assai drammatico di disadattamento scolastico che molto spesso prelude all'abbandono degli studi.

Comportamento scolastico e classe sociale

Grazie alle informazioni sulla condizione professionale, sulla posizione nella professione e sul titolo di studio dei genitori degli alunni intervistati, è ora possibile estendere la stima dei parametri scolastici fin qui analizzati, ai vari sottogruppi riconducibili ai diversi livelli socio-economici delle famiglie da cui gli alunni provengono.

Le tabelle 11, 12, 13, 14 riportano dettagliatamente, per i cinque anni di corso, per i quattro gruppi sociali considerati - classe alta, media, bassa, non professionale - i parametri relativi a provenienza - esito - destinazione fin qui considerati in forma globale e distinti per sesso.

In linea generale, da un'attenta analisi dei dati, sono possibili le seguenti considerazioni:

- 1. i tassi di iscrizione alla scuola post-obbligo risentono ancora sensibilmente del livello sociale di provenienza degli alunni, privilegiando coloro che provengono dalle classi medie e inferiori. Anche i tassi di regolarità, specialmente nel primo triennio di studi superiori, risentono dell'origine sociale degli allievi. Viceversa, i tassi di ripetenza non sembrano più essere influenzati in modo sensibile dalla classe sociale di appartenenza; ciò avviene in modo analogo per i tassi di promozione e di bocciatura, che solo nella prima classe tendono, sia pure lievemente, a sanzionare le classi meno abbienti.*
- 2. I parametri che meglio evidenziano il permanere dell'elevato grado di selettività del sistema scolastico post-obbligo della provincia di Trento sono tre: innanzitutto il tasso globale di abbandono degli studi nel passaggio dalla 3^a media alla 1^a superiore e la gerarchizzazione dei percorsi scolastici post-obbligo secondo i noti livelli di prestigio dei*

*diversi indirizzi scolastici*¹. Un secondo indicatore di selettività è rappresentato dal tasso di abbandono dei bocciati, che per i primi anni del ciclo risulta quasi doppio per gli alunni di classe bassa (84.7% dei bocciati abbandona) rispetto a quelli di classe alta (42.9% dei bocciati abbandona). Infine il tasso di uscita con titolo (graduation) che alla fine del secondo e del terzo anno di corso è quasi cinque volte superiore per le classi basse (47.3% di uscite alla fine della terza) rispetto alle classi alte (10.5% di uscite con titolo alla fine della terza).

- 3. La selettività del sistema scolastico, pertanto, deve trovare i suoi punti di misurazione proprio nei punti di cerniera fra sistema scolastico dell'obbligo e sistema scolastico post-obbligo, nel primo anno di tale ciclo, ed in particolare nelle modalità di abbandono precoce, con o senza titolo, con o senza bocciatura esplicita, degli studi. Un ulteriore punto importante, emerso nettamente in questa ricerca, riguarda la mobilità degli studenti bocciati ed irregolari, che cambiano frequentemente sede scolastica ed indirizzo di studi, con conseguenti disagi, sradicamenti e disadattamenti personali, culturali e sociali, oltre che scolastici.*

¹ Nel già citato articolo di G. Chiari, A. de Lillo.

Tabella 1 - Stima dei flussi di uscita (per esito e destinazione) dal sistema scolastico dell'obbligo della provincia di Trento, alla fine dell'anno scolastico 1973/74 dei frequentanti la classe III media.

	Valori assoluti					
	Totale	Abbandonano	Proseguono nella		Formaz. Profess.	Fuori Prov.
			III media	Secondaria		
Licenziati	6390	1166	-	3465	1699	60
Bocciati	273	60	213	-	-	-
Totale	6663	1226	213	3465	1699	60
Valori percentuali						
Licenziati	95,9	17,5	-	52,0	25,5	0,9
Bocciati	4,1	0,9	3,2	-	-	-
Totale	100,0	18,4	3,2	52,0	25,5	0,9

Tabella 2 - Stima dei flussi di provenienza degli alunni iscritti al primo anno della scuola secondaria post-obbligo in provincia di Trento nell'anno scolastico 1974/75

	Totale					
	Valori assoluti	Valori percentuali				
Regolari (prov. TN)	4431	84,3				
Ripetenti	519	9,8				
Rientri	125	2,4				
Non identificati	183	3,5				
Totale	5258	100,0				
	Sesso		Livello Sociale			
	Maschi	Femmine	Alto	Medio	Basso	Non Prof.
Regolari (prov. TN)	82,1	86,7	89,1	85,3	83,7	84,4
Ripetenti	12,4	7,0	2,9	10,2	9,0	9,4
Rientri	1,6	3,3				
Non identificati	3,9	3,5				
Totale	100,0	100,0	100,0	100,0	100,0	100,0

NB: L'eventuale non corrispondenza fra i dati delle diverse tabelle pubblicate è dovuto al fatto che alcuni allievi non hanno risposto alle domande relative alle variabili di base.

Tabella 3 - Alunni iscritti alla scuola secondaria post-obbligo della Provincia di Trento, ai vari anni di corso, per provenienza, esito e destinazione

	Valori assoluti					Totale
	I 1974/75	II 1975/76	III 1976/77	IV 1977/78	V 1978/79	
Iscritti	5258	4698	3676	2240	1930	
Provenienza:						
Regolari	4431	4446	3500	2114	1777	
Ripetenti	519	236	176	91	57	
Rientri	125	9	-	7	3	144
N.I. (1)	183	7	-	28	93	311
Destinazione:						
Licenziati	42	654	1174	289	1825	3984 (2)
Abbandoni	497	304	163	41	-	1055
Proseguono	4446	3500	2114	1777	-	
Ripetono	266	223	202	128	102	102
N.I.	7	17	23	42	3	89
Esito:						
Promossi	4583	4235	3321	2071	1825	
Bocciati	670	454	346	164	102	
N.I.	5	9	9	5	3	28

(1) N.I. = non identificato

(2) Il tasso lordo di realizzazione (regolari più 1 anno di ritardo) è pari al 75.77%

Tabella 3 bis - Sistema scolastico post-obbligo della provincia di Trento: tassi netti di proseguimento e di permanenza. Dati globali.

Tassi netti di:	I - II	II - III	III - IV	IV - V
Proseguimento regolare	84,6	74,5	57,5	79,3
Permanenza	94,4	79,5	62,3	83,4
Tavola di:				
Regolarità	84,6	63,0	36,2	28,7
Permanenza	94,4	75,0	46,7	39,0

Tabella 3 ter - Sistema scolastico post-obbligo della provincia di Trento.
Tassi netti di proseguimento regolare e di permanenza: andamenti per sesso

	Maschi					Femmine				
	I	II	III	IV	V	I	II	III	IV	V
Tassi netti di:										
Proseguimento regolare	81.4	72.2	54.0	86.9	-	88.2	69.4	62.0	70.9	-
Permanenza	88.2	84.9	60.9	95.5	*	91.3	73.0	65.7	73.4	*
Tavole di:										
Regolarità	100.0	81.4	58.8	31.7	27.8	100.0	88.2	61.2	37.9	26.9
Permanenza	100.0	88.2	74.9	45.6	43.6	100.0	91.3	66.6	43.8	32.1
Tassi di:										
Ripetenza	6.8	5.7	6.9	8.6	*	3.1	3.6	3.7	2.5	*
Abbandono	11.6	7.5	4.8	1.4	*	7.0	5.4	4.0	2.4	*

Tabella 4 - Alunni iscritti alla scuola secondaria post-obbligo della Provincia di Trento, ai vari anni di corso, per provenienza, esito e destinazione

	Valori percentuali				
	I	II	III	IV	V
Iscritti (N)	5258	4698	3676	2240	1930
Provenienza:					
Regolari	84.3	94.6	95.2	94.4	92.1
Ripetenti	9.8	5.0	4.8	4.0	2.9
Rientri	2.4	0.2	-	0.3	0.2
N.I.	3.5	0.1	-	1.2	4.8
Destinazione:					
Proseguono	84.6	74.5	57.5	79.3	-
Licenziati	0.8	13.9	31.9	12.9	94.6
Abbandoni	9.4	6.5	4.4	1.8	-
Ripetono	5.0	4.7	5.5	5.7	-
N.I.	0.2	0.6	0.8	2.1	5.4
Esito:					
Promossi	87.1	90.1	90.3	92.4	94.5
Bocciati	12.7	9.7	9.4	7.3	5.3

Tabella 5 - Dati complessivi di comportamento scolastico post-obbligo in Provincia di Trento: andamenti per maschi e femmine

	Valori assoluti									
	B Maschi					C Femmine				
	I	II	III	IV	V	I	II	III	IV	V
Iscritti	2819	2456	2064	1178	1097	2439	2242	1612	1062	833
Provenienza:										
Regolari	2316	2294	1945	1114	1024	2115	2152	1555	1000	753
Ripetenti	349	155	126	50	44	170	81	50	41	19
Rientri	45	1	-	5	3	80	8	-	2	-
N.I.	109	6	-	9	26	74	1	-	19	67
Destinazione:										
Proseguono	2294	1945	1114	1024	-	2152	1555	1000	753	-
Licenziati	2	174	694	35	1021	40	480	480	254	804
Abbandoni	326	185	99	16	*	171	121	64	25	*
Ripetono	191	141	143	101	*	75	82	59	27	*
N.I.	6	11	14	2	76	1	4	9	3	29
Esito:										
Promossi	2356	2157	1824	1063	1021	2227	2078	1497	1008	804
Bocciati	459	292	234	113	74	211	162	112	51	28
N.I.	4	7	6	2	2	1	2	3	3	1
% bocciati che ripetono	41.6	48.2	61.1	89.4	*	35.5	50.6	52.7	52.9	*

Tabella 6 - Dati complessivi di comportamento scolastico post-obbligo: andamenti per maschi e femmine

	Valori percentuali									
	Maschi					Femmine				
	I	II	III	IV	V	I	II	III	IV	V
Iscritti										
Provenienza:										
Regolari	82.1	93.4	94.2	94.6	93.3	86.7	96.0	96.5	94.2	90.4
Ripetenti	12.4	6.3	6.1	4.2	4.0	7.0	3.6	3.1	3.9	2.3
Rientri	1.6	0.0	-	0.4	0.3	3.3	0.4	-	0.2	-
N.I.	3.9	0.2	-	0.8	2.4	3.0	0.0	-	1.8	8.0
Destinazione:										
Proseguono	81.4	79.2	54.0	86.9	-	88.2	69.4	62.0	70.9	-
Licenziati	0.1	7.1	33.6	3.0	93.1	1.6	21.4	29.8	23.9	96.5
Abbandoni	11.6	7.5	4.8	1.4	*	7.0	5.4	4.0	2.4	*
Ripetono	6.8	5.7	6.9	8.6	*	3.1	3.6	3.7	2.5	*
N.I.	0.2	0.4	0.7	6.9	0.0	0.2	0.6	0.3	3.5	
Esito:										
Promossi	83.6	87.8	88.4	90.2	93.1	91.3	92.7	92.9	94.9	96.5
Bocciati	16.3	11.9	11.3	9.6	6.7	8.6	7.2	6.9	4.8	3.4
N.I.	0.1	0.3	0.3	0.2	0.2	0.0	0.1	0.2	0.3	0.1

Figura 1 - Andamento dei tassi di proseguimento, bocciatura e abbandono nel sistema scolastico post-obbligo della provincia di Trento. Differenza maschi-femmine.

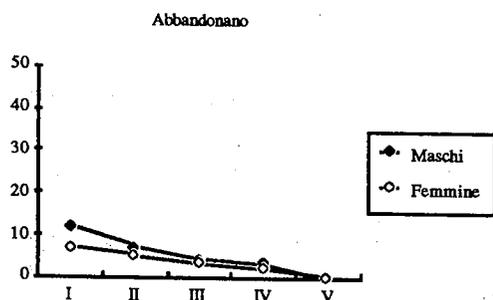
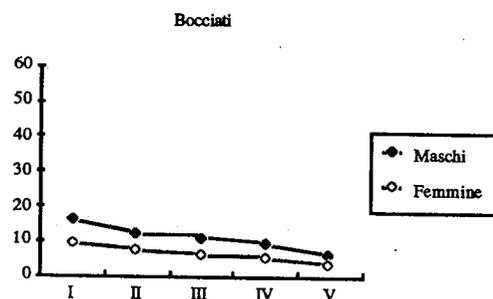
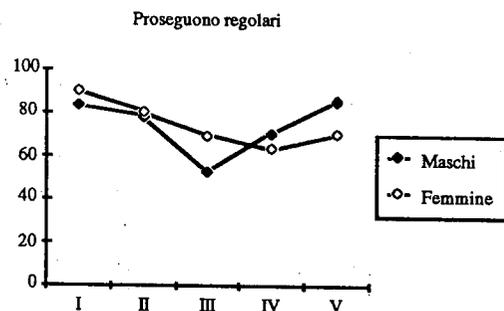


Tabella 7 - Tassi di promozione e di bocciature nel sistema scolastico post-obbligo della provincia di Trento: dati globali e per sesso

	Totale				
	I	II	III	IV	V
% Promossi	87.1	90.1	90.3	92.4	94.5
% Bocciati	12.7	9.7	9.4	7.3	5.3
Ripetono	5.0	4.7	5.5	5.7	5.3 *
Abbandonano	7.6	4.7	3.5	1.6	-
N.I.	0.1	0.3	0.4	-	-
	Maschi				
% Promossi	83.6	87.8	88.4	90.2	93.1
% Bocciati	16.3	11.9	11.3	9.6	6.7
Ripetono	6.7	5.7	6.9	8.5	*
Abbandonano	9.4	5.9	4.0	1.0	-
N.I.	0.1	0.3	0.3	0.2	0.2
	Femmine				
% Promossi	91.3	92.7	92.9	94.9	96.5
% Bocciati	8.6	7.2	6.9	4.8	3.4
Ripetono	3.0	3.6	3.7	2.5	*
Abbandonano	5.6	3.4	2.9	2.5	*
N.I.	0.1	0.1	0.2	0.3	0.1

Figura 2 - Tassi di bocciatura nelle varie classi della scuola post-obbligo: differenze maschi-femmine

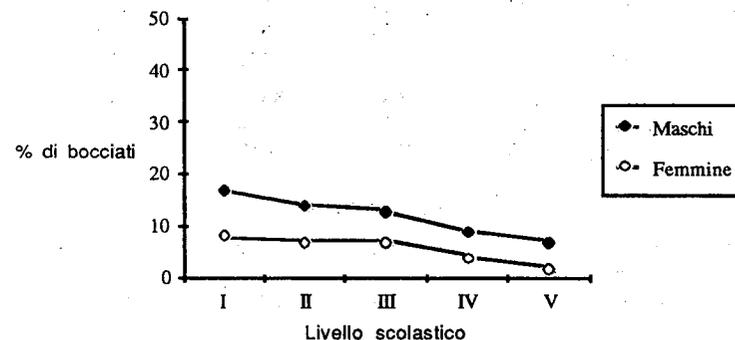


Figura 3 - % di bocciati che ripetono

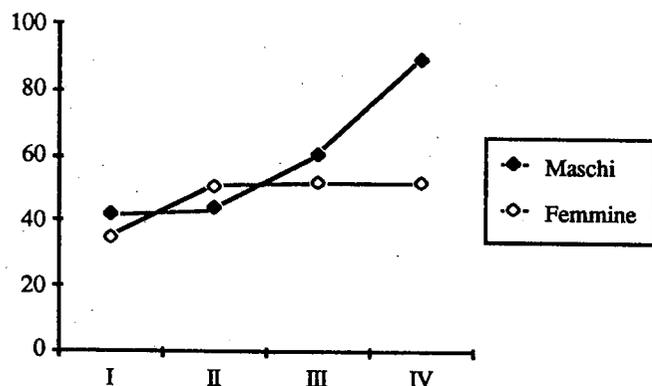


Tabella 8 - Diplomatici e qualificati della scuola post-obbligo: dati globali e per sesso

		Valori assoluti				
		I	II	III	IV	V
Licenziati:	Totale	42	654	1174	289	1825
	Maschi	2	174	694	35	1021
	Femmine	40	480	480	254	804
Iscritti:	Totale	5258	4698	3676	2240	1930
	Maschi	2819	2456	2064	1178	1097
	Femmine	2439	2242	1612	1062	833
		Valori percentuali				
% licenziati:	Totale	0.80	13.9	31.9	12.9	94.5
	Maschi	0.07	7.0	33.6	2.9	93.0
	Femmine	1.6	21.4	29.8	23.9	96.5

Tabella 9 - Tipi di percorso irregolare

Hanno cambiato:	Valori assoluti					Totale
	I	II	III	IV	V	
Istituto:						
Regolari	-	18	111	57	4	190
Ripetenti	61	6	7	-	-	74
Istituto e tipo di studi:						
Regolari	5	10	16	57	3	91
Ripetenti	146	24	4	-	-	174
Istituto e provincia:						
Regolari	13	13	9	1	-	36
Ripetenti	6	3	1	-	-	10
Istituto, tipo di studi e provincia:						
Regolari	123	2	9	11	5	150
Ripetenti	10	3	2	1	-	16
Rientri da abbandono	125	9	4	7	3	148
Totale	489	88	163	134	15	889 *
Totale iscritti	5258	4698	3676	2240	1930	17802

Tabella 10 - Tipi di percorso irregolare

Hanno cambiato:	Valori percentuali					Totale	%
	I	II	III	IV	V		
Istituto:							
Regolari	-	0.4	3.0	2.5	0.2	190	21.4
Ripetenti	1.2	0.1	0.2	-	-	74	8.3
Istituto e tipo di studi:							
Regolari	0.1	0.2	0.4	2.5	0.2	91	10.2
Ripetenti	2.8	0.5	0.1	-	-	174	19.6
Istituto e provincia:							
Regolari	0.2	0.3	0.2	0.0	-	36	4.0
Ripetenti	0.1	0.1	0.0	-	-	10	1.1
Istituto, tipo di studi e provincia:							
Regolari	2.3	0.0	0.2	0.5	0.3	150	16.9
Ripetenti	0.2	0.1	0.1	0.0	-	16	1.8
Rientri da abbandono	2.4	0.2	0.1	0.3	0.2	148	16.6
Totale	9.3	1.9	4.4	6.0	0.8	889	100.0
Totale iscritti	5258	4698	3676	2240	1930	17802	
Regolari	4431	4446	3500	2114	1777	16268	
Ripetenti	519	236	176	91	57	1079	
Rientri	125	9	-	7	3	144	
Non ident.	183	7	-	28	93	311	

Tabella 11 - Comportamento scolastico per classe sociale

	Valori assoluti				
	I	II	III	IV	V
Iscritti:					
Classe alta	412	402	359	278	250
Classe media	2051	1834	1477	986	832
Classe bassa	1956	1705	1272	602	499
Classe non prof.	476	425	317	158	138
Totale	4895	4366	3425	2024	1719
Provenienza regolari:					
Classe alta	367	335	278	243	237 *
Classe media	1749	1426	969	788	798 *
Classe bassa	1637	1228	588	478	470 *
Classe non prof.	402	299	148	128	135 *
Totale	4155	3288	1983	1637	1640
Ripetenti:					
Classe alta	12	33	25	3	1
Classe media	209	83	53	10	13
Classe bassa	177	65	49	5	7
Classe non prof.	45	24	19	6	6
Totale	443	205	146	24	27
	93.1	92.9	93.2	90.4	89.1
Bias:	5258	100.0			
n.i.	4895	93.1	6.9	1^ bias	
	297	5.6	5.6	2^ bias	
			12.5	bias tot.	

* licenziati

Tabella 12 - Comportamento scolastico per classe sociale

	Valori percentuali				
	I	II	III	IV	V
Iscritti:					
Classe alta	8.4	9.2	10.5	13.7	14.5
Classe media	41.9	42.0	43.1	48.7	48.4
Classe bassa	40.0	39.0	37.1	29.7	29.0
Classe non prof.	9.7	9.7	9.3	7.8	8.0
Regolari:					
Classe alta	89.1	83.3	77.4	87.4	94.8
Classe media	85.3	77.7	65.6	79.9	95.9
Classe bassa	83.7	72.0	46.2	79.4	94.2
Classe non prof.	84.4	70.3	46.7	81.0	97.8
Totale	84.9	75.3	57.9	80.9	95.4
Ripetenti:					
Classe alta	2.9	8.2	7.0	1.1	0.4
Classe media	10.2	4.5	3.6	1.0	1.6
Classe bassa	9.0	3.8	3.8	0.8	1.4
Classe non prof.	9.4	5.6	6.0	3.8	4.3
Totale	8.1	4.7	4.3	1.2	1.6

Tabella 13 - Comportamento scolastico per classe sociale

	Valori assoluti				
	I	II	III	IV	V
Bocciati:					
Classe alta	42	41	43	22	13
Classe media	256	171	140	71	34
Classe bassa	256	154	97	40	27
Classe non prof.	59	17	33	7	3
Totale	613	383	313	140	77
Abbandonano:					
Classe alta	18	20	17	5	*
Classe media	184	102	63	14	*
Classe bassa	199	115	51	14	*
Classe non prof.	50	32	16	2	*
Totale	451	269	147	35	*
Promossi:					
Classe alta	370	359	314	257	237
Classe media	1790	1661	1334	912	798
Classe bassa	1699	1548	1171	561	470
Classe non prof.	417	379	283	151	135
Totale	4276	3947	3102	1881	1640
Licenziati:					
Classe alta	-	18	33	14	237
Classe media	11	207	354	121	798
Classe bassa	24	290	567	83	470
Classe non prof.	5	73	134	23	135
Totale	40	588	1088	241	1640

Tabella 14 - Comportamento scolastico per classe sociale

	Valori percentuali				
	I	II	III	IV	V
Bocciati:					
Classe alta	10.2	10.2	12.0	7.9	5.2
Classe media	12.5	9.3	9.5	7.2	1.6
Classe bassa	13.1	9.0	7.6	6.6	5.4
Classe non prof.	12.4	4.0	10.4	4.4	2.2
Totale	12.5	8.8	9.1	6.9	4.5
Abbandonano:					
Classe alta	42.9	48.8	39.5	22.7	*
Classe media	71.9	59.6	45.0	19.7	*
Classe bassa	77.7	74.7	52.6	35.0	*
Classe non prof.	84.7	*	48.5	28.6	*
Totale	73.6	70.2	47.0	25.0	*
Promossi:					
Classe alta	89.8	89.3	87.5	92.4	94.8
Classe media	87.3	90.6	90.3	92.5	95.9
Classe bassa	86.9	90.8	92.1	93.2	94.2
Classe non prof.	87.6	89.2	89.3	95.6	97.8
Totale	87.3	90.4	90.6	92.9	95.4
Diplomati o qualificati:					
Classe alta	-	5.0	10.5	5.4	100.0
Classe media	0.6	12.5	26.5	13.3	100.0
Classe bassa	1.4	18.7	48.4	14.8	100.0
Classe non prof.	1.2	19.8	47.3	15.2	100.0
Totale	0.9	14.9	35.1	12.8	100.0

Tabella 15 - Alcuni iscritti nel sistema scolastico trentino dalla III media alla V superiore: andamenti per sesso, professione e titolo di studio del padre, zona di residenza e per provenienza, esito e destinazione.

Valori assoluti	1973/74	1974/75	1975/76	1976/77	1977/78	1978/79
	III Media	I Superiore	II Superiore	III Superiore	IV Superiore	V Superiore
Sesso:						
M	3476	2816	2449	2056	1191	1095
F	3451	2437	2238	1607	1044	831
Totale	6927	5253	4687	3663	2235	1926
N. indici	100	75.8	67.7	52.9	32.3	27.8
		100.0	89.2	69.7	42.5	36.7
Professione del padre:						
Alta	420	412	400	357	277	250
Media	2490	2047	1832	1474	983	833
Bassa	3488	2429	2129	1587	759	635
Totale	6398	4888	4361	3418	2019	1718
Titolo di studio del padre:						
Alto	678	697	629	563	454	415
Medio	1564	1406	1299	1060	695	587
Basso	4186	2853	2495	1844	902	746
Totale	6428	4956	4423	3467	2051	1748
Zona di residenza:						
Alta	3493	2895	2715	2191	1381	1222
Media	1229	833	770	545	274	242
Bassa	1713	1088	992	731	357	286
Totale	6435	4816	4477	4467	2012	1750
Provenienza:						
Da promozione	6543	4420	4164	3279	2000	1698
Da ripetenza	189	471	205	146	24	21
Totale	6732	4891	4368	3425	2024	1719
Esito:						
Promosso	6663	4583	4234	3321	2072	1824
Respinto	265	670	453	342	163	102
Totale	6928	5253	4687	3663	2235	1926
Destinazione:						
Continua:						
- stesso Comune	1911					
- stesso Distretto	1572					
- Fuori Distretto	1155					
- Fuori Provincia	90					
Totale	4728					
Abbandona	2005					
Totale	6733					

Tabella 15 (segue) -

Valori percentuali	1973/74	1974/75	1975/76	1976/77	1977/78	1978/79
	III Media	I Superiore	II Superiore	III Superiore	IV Superiore	V Superiore
Sesso:						
M	50.2	53.6	52.2	56.1	53.3	56.9
F	49.8	46.4	47.8	43.9	46.7	43.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0
Professione del padre:						
Alta	6.6	8.4	9.2	10.4	13.7	14.6
Media	38.9	41.9	42.0	43.1	48.7	48.5
Bassa	54.5	49.7	48.8	46.5	37.6	36.9
Totale	100.0	100.0	100.0	100.0	100.0	100.0
Titolo di studio del padre:						
Alto	10.5	14.1	14.2	16.2	22.1	23.7
Medio	24.3	28.4	29.4	30.6	33.9	33.6
Basso	65.2	57.5	56.4	53.2	44.0	42.7
Totale	100.0	100.0	100.0	100.0	100.0	100.0
Zona di Residenza:						
Alta	54.3	60.1	60.6	49.0	68.6	69.8
Media	19.1	17.3	17.2	12.1	13.6	13.8
Bassa	26.6	22.6	22.2	38.9	17.8	16.4
Totale	100.0	100.0	100.0	100.0	100.0	100.0
Provenienza:						
Da promozione	97.2	90.4	95.3	95.7	98.8	98.8
Da ripetenza	2.8	9.6	4.7	4.3	1.2	1.2
Totale	100.0	100.0	100.0	100.0	100.0	100.0
Esito:						
Promosso	96.2	87.2	90.3	90.7	92.7	94.7
Respinto	3.8	12.8	9.7	9.3	7.3	5.3
Totale	100.0	100.0	100.0	100.0	100.0	100.0
Destinazione:						
Continua:						
- stesso Comune	28.4					
- stesso Distretto	23.3					
- Fuori Distretto	17.1					
- Fuori Provincia	1.3					
Totale	70.2					
Abbandona	29.8					
Totale	100.0					

Tabella 16a - Provenienza degli alunni secondo la professione del padre: I-II-III-IV-V superiore. Valori assoluti.

	Professione del padre											
	TOTALE			ALTA			MEDIA			BASSA		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
I^A Superiore												
Regolari normali	4184	2186	1998	360	195	165	1731	889	842	2093	1102	991
Regolari atipici	127	68	59	9	7	2	68	32	36	50	29	21
Ripetenti	471	315	156	42	27	15	209	135	74	220	153	67
Rientri	104	35	69	1	0	1	40	17	23	63	18	45
Non identificato	5	2	3	0	0	0	2	1	1	3	1	2
N	4891	2606	2285	412	229	183	2050	1074	976	2429	1303	1126
II^A Superiore												
Regolari normali	4113	2104	2009	362	193	169	1731	867	864	2020	1044	976
Regolari atipici	38	23	15	4	2	2	16	10	6	18	11	7
Ripetenti	205	133	72	33	20	13	83	51	32	89	62	27
Rientri	6	1	4	2	0	1	1	0	1	3	1	2
Non identificato	6	1	3	1	0	1	3	0	2	2	1	0
N	4368	2262	2103	402	215	186	1834	928	905	2132	1119	1012
III^A Superiore												
Regolari normali	3149	1699	1449	324	174	150	1364	702	661	1961	823	638
Regolari atipici	126	97	29	10	7	3	59	41	18	57	49	8
Ripetenti	146	107	38	25	18	6	53	38	15	68	51	17
Rientri	3	1	2	0	0	0	0	0	0	3	1	2
Non identificato	1	1	0	0	0	0	1	1	0	0	0	0
N	3425	1905	1518	359	199	159	1477	782	694	1589	924	665
IV^A Superiore												
Regolari normali	1925	1006	919	270	145	125	949	472	477	706	389	317
Regolari atipici	71	35	36	5	3	2	27	13	14	39	19	20
Ripetenti	24	15	9	3	1	2	10	4	6	11	10	1
Rientri	3	3	0	0	0	0	0	0	0	1	1	0
Non identificato	1	1	0	0	0	0	0	0	0	1	1	0
N	2024	1060	964	278	149	129	986	489	497	760	422	317
V^A Superiore												
Regolari normali	1689	931	758	246	131	115	825	423	402	618	377	241
Regolari atipici	7	2	5	2	0	2	1	0	1	4	2	2
Ripetenti	21	19	1	1	1	0	7	7	0	13	11	1
Rientri												
Non identificati	2	2	0	1	1	0	0	0	0	1	1	0
N	1719	954	764	250	133	117	833	430	403	636	391	244

Tabella 16b - Provenienza degli alunni secondo la professione del padre: I - II - III - IV - V superiore. Valori percentuali.

	Professione del padre											
	TOTALE			ALTA			MEDIA			BASSA		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
I^A Superiore												
Regolari normali	87.4	85.2	90.2	84.4	82.8	86.3	86.2	84.6	88.0	85.5	83.9	87.4
Regolari atipici	2.2	3.1	1.1	3.3	3.0	3.7	2.1	3.1	1.9	2.6	2.6	2.6
Ripetenti	10.2	11.8	8.2	10.2	12.6	7.6	9.1	11.7	6.0	9.6	12.1	6.8
Rientri	0.2	0.0	0.5	2.0	1.6	2.4	2.6	1.4	4.0	2.1	1.3	3.0
Non identificato	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	412	229	183	2050	1074	976	2429	1303	1126	4891	2606	2285
II^A Superiore												
Regolari normali	90.0	89.8	90.9	94.4	93.4	95.5	94.7	93.3	96.4	94.2	93.0	95.5
Regolari atipici	1.0	0.9	1.1	0.9	1.1	0.7	0.8	1.0	0.7	0.9	1.0	0.7
Ripetenti	8.2	9.3	7.0	4.5	5.5	3.5	4.2	5.5	2.7	4.7	5.9	3.4
Rientri	0.5	0.0	0.5	0.1	0.0	0.1	0.1	0.1	0.2	0.1	0.0	0.2
Non identificato	0.2	0.0	0.5	0.2	0.0	0.2	0.1	0.1	0.0	0.1	0.0	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	402	215	186	1834	928	905	2132	1119	1012	4368	2262	2103
III^A Superiore												
Regolari normali	90.3	87.4	94.3	92.3	89.8	95.2	91.9	89.1	95.9	91.9	89.2	95.5
Regolari atipici	2.8	3.5	1.9	4.0	5.2	2.6	3.6	5.3	1.2	3.7	5.1	1.9
Ripetenti	7.0	9.0	3.8	3.6	4.9	2.2	4.3	5.5	2.6	4.3	5.6	2.5
Rientri	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.3	0.1	0.1	0.1
Non identificato	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	359	199	159	1477	782	694	1589	924	665	3925	1905	1518
IV^A Superiore												
Regolari normali	97.1	97.3	96.9	96.2	96.5	96.0	92.9	92.2	93.8	95.1	94.9	95.3
Regolari atipici	1.8	2.0	1.6	2.7	2.7	2.8	5.1	4.5	5.9	3.5	3.3	3.7
Ripetenti	1.1	0.7	1.6	1.0	0.8	1.2	1.4	2.4	0.3	1.2	1.4	0.9
Rientri	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.7	0.0	0.1	0.3	0.0
Non identificato	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.0
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	278	149	129	986	489	497	760	422	338	2024	1060	964
V^A Superiore												
Regolari normali	98.4	98.5	98.3	99.0	98.9	99.8	97.2	96.4	98.8	98.3	97.1	99.2
Regolari atipici	0.8	0.0	1.7	0.1	0.0	0.2	0.6	0.5	0.8	0.4	0.2	0.7
Ripetenti	0.4	0.8	0.0	0.8	1.6	0.0	2.0	2.8	0.4	1.2	2.0	0.1
Rientri												
Non identificato	0.4	0.8	0.0	0.0	0.0	0.0	0.2	0.3	0.0	0.1	0.2	0.0
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	250	133	117	833	430	403	636	391	244	1719	954	764

Tabella 17a - Provenienza degli alunni secondo il titolo di studio del padre: I - II - III - IV - V superiore. Valori assoluti.

I ^A Superiore	Titolo di studio del padre											
	TOTALE			ALTA			MEDIA			BASSA		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
Regolari normali	4242	2214	2028	587	301	286	1188	599	589	2467	1314	1153
Regolari atipici	128	69	59	23	16	7	43	18	25	62	35	27
Ripetenti	480	320	160	82	56	26	154	98	56	284	166	78
Rientri	103	34	69	5	2	3	21	8	13	77	24	53
Non identificato	6	3	3	0	0	0	1	0	1	5	3	2
N	4959	2640	2319	697	375	322	1407	723	684	1855	1542	1313
II ^A Superiore												
Regolari normali	4163	2130	2033	596	310	286	1197	576	621	2370	1244	1126
Regolari atipici	42	25	17	5	3	2	13	8	5	24	74	10
Ripetenti	214	140	74	28	15	13	87	59	28	99	66	33
Rientri	6	1	5	1	0	1	3	0	3	2	1	1
Non identificato	4	1	3	0	0	0	1	1	0	3	0	3
N	4429	2297	2132	630	328	302	1302	644	657	2499	1325	1173
III ^A Superiore												
Regolari normali	3183	1716	1467	529	268	261	971	491	480	1683	957	726
Regolari atipici	130	103	27	12	6	6	34	29	5	84	68	16
Ripetenti	154	111	43	23	15	8	54	42	12	77	54	23
Rientri	4	2	2	0	0	0	1	1	0	3	1	2
Non identificato	1	1	0	0	0	0	0	0	0	1	1	0
N	3472	1933	1539	564	289	275	1060	563	497	1849	1081	767
IV ^A Superiore												
Regolari normali	1953	1025	928	444	222	222	666	347	319	843	456	387
Regolari atipici	72	37	35	6	5	1	20	8	12	46	24	22
Ripetenti	17	18	9	5	2	3	10	4	6	12	12	0
Rientri	3	3	0	0	0	0	1	1	0	2	2	0
Non identificato	1	1	0	0	0	0	0	0	0	1	1	0
N	2056	1084	972	455	229	226	697	360	337	904	495	409
V ^A Superiore												
Regolari normali	1715	950	765	411	207	204	577	313	264	727	430	297
Regolari atipici	7	2	5	1	0	1	2	0	2	4	2	2
Ripetenti	23	22	1	1	1	0	8	7	1	14	14	0
Rientri	1	1	0	0	0	0	0	0	0	1	1	0
Non identificato	2	2	0	1	1	0	0	0	0	1	1	0
N	1748	977	771	415	209	205	587	320	267	747	448	299

Tabella 17b - Provenienza degli alunni secondo il titolo di studio del padre: I - II - III - IV - V superiore. Valori percentuali.

I ^A Superiore	Titolo di studio del padre											
	ALTO			MEDIO			BASSO			TOTALE		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
Regolari normali	84.2	80.3	88.8	84.4	82.8	86.1	86.4	85.2	87.8	85.7	83.9	87.5
Regolari atipici	3.3	4.3	2.2	3.1	2.5	3.7	2.1	2.3	2.1	2.5	2.6	2.5
Ripetenti	11.8	14.9	8.1	10.9	13.6	8.2	8.5	10.8	5.9	9.7	12.1	6.9
Rientri	0.7	0.5	0.9	1.5	1.1	1.9	2.7	1.6	4.0	2.0	1.3	3.0
Non identificato	0.0	0.0	0.0	0.1	0.0	0.1	0.2	0.2	0.2	0.1	0.1	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	697	375	322	1407	723	684	2855	1542	1313	4959	2640	2319
II ^A Superiore												
Regolari normali	94.6	94.5	94.7	91.9	89.4	94.5	94.8	93.9	96.0	94.0	92.7	95.4
Regolari atipici	0.8	0.9	0.7	1.0	1.2	0.8	1.0	1.1	0.9	1.0	1.1	0.8
Ripetenti	4.4	4.6	4.3	6.7	9.2	4.3	4.0	5.0	2.8	4.8	6.1	3.5
Rientri	0.2	0.0	0.3	0.2	0.0	0.5	0.1	0.1	0.1	0.1	0.0	0.2
Non identificato	0.0	0.0	0.0	0.2	0.2	0.0	0.2	0.0	0.3	0.1	0.0	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	630	328	302	1302	644	657	2499	1325	1173	4429	2297	2132
III ^A Superiore												
Regolari normali	93.6	92.7	94.4	91.6	87.2	96.6	91.1	88.5	94.7	91.7	88.8	95.3
Regolari atipici	2.1	2.1	2.2	3.2	5.2	1.0	4.5	6.3	2.1	3.7	5.3	1.8
Ripetenti	4.2	5.2	2.9	5.1	7.5	2.4	4.2	5.0	3.0	4.4	5.7	2.8
Rientri	0.0	0.0	0.0	0.1	0.2	0.0	0.2	0.1	0.3	0.1	0.1	0.1
Non identificato	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.1	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	564	289	275	1060	563	497	1849	1081	767	3472	1933	1539
IV ^A Superiore												
Regolari normali	97.6	96.9	98.2	95.6	96.4	94.7	93.3	92.1	94.6	95.0	94.6	95.5
Regolari atipici	1.3	2.2	0.4	2.9	2.2	3.6	5.1	4.8	5.4	3.5	3.4	3.6
Ripetenti	1.1	0.9	1.3	1.4	1.1	1.8	1.3	2.4	0.0	1.3	1.7	0.9
Rientri	0.0	0.0	0.0	0.1	0.3	0.0	0.2	0.4	0.0	0.1	0.3	0.0
Non identificato	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.1	0.1	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	455	229	226	697	360	337	904	495	409	2056	1084	972
V ^A Superiore												
Regolari normali	99.0	99.0	99.5	98.3	97.8	98.9	97.3	96.0	99.3	98.1	97.2	99.2
Regolari atipici	0.2	0.0	0.5	0.3	0.0	0.7	0.5	0.4	0.7	0.4	0.2	0.6
Ripetenti	0.5	0.5	0.0	1.4	2.2	0.4	1.9	3.1	0.0	1.3	2.3	0.1
Rientri	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.0	0.1	0.0
Non identificati	0.2	0.5	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.2	0.2	0.0
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	415	209	205	587	320	267	747	448	299	1748	977	771

Tabella 18a - Provenienza degli alunni secondo la zona di residenza: I - II - III - IV - V superiore. Valori assoluti.

I ^A Superiore	Zona di residenza											
	TOTALE			ALTA			MEDIA			BASSA		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
Regolari normali	4160	2186	1974	2493	1279	1214	728	379	349	539	528	411
Regolari atipici	53	28	25	32	20	12	8	4	4	13	4	9
Ripetenti	490	332	158	323	214	109	71	50	21	96	68	28
Rientri	108	36	72	43	14	29	26	8	18	39	14	25
Non identificato	5	3	2	4	3	1	0	0	0	1	0	1
N	4816	2585	2231	2895	1530	1365	833	441	392	1088	614	479
II ^A Superiore												
Regolari normali	4217	2171	2046	2548	1291	1257	734	380	354	935	500	435
Regolari atipici	39	23	16	26	13	13	6	6	0	7	4	3
Ripetenti	216	140	76	138	92	46	31	18	13	47	30	17
Rientri	8	1	6	5	0	5	0	0	0	3	1	2
Non identificato	6	1	3	4	0	4	1	0	0	1	1	0
N	4486	2336	2147	2721	1396	1325	772	404	368	993	536	457
III ^A Superiore												
Regolari normali	3194	1734	1460	2044	1077	967	488	271	217	662	386	276
Regolari atipici	132	104	28	57	43	14	36	30	6	39	31	8
Ripetenti	147	105	42	93	64	29	22	17	5	32	24	8
Rientri	2	1	1	2	1	1	0	0	0	0	0	0
Non identificato	3	3	0	3	3	0	0	0	0	0	0	0
N	3478	1947	1531	2199	1188	1011	546	318	228	733	441	292
IV ^A Superiore												
Regolari normali	1867	963	904	1307	670	637	238	120	118	322	173	149
Regolari atipici	66	31	35	21	9	12	28	13	15	17	9	8
Ripetenti												
Rientri												
Non identificato												
N	1933	994	939	1328	679	649	266	133	133	339	182	157
V ^A Superiore												
Regolari normali	1720	961	759	1201	652	549	235	137	98	284	172	112
Regolari atipici	7	2	5	5	1	4	1	0	1	1	1	0
Ripetenti	23	11	2	15	14	1	7	7	0	1	0	1
Rientri	1	1	0	1	1	0	0	0	0	0	0	0
Non identificato	1	1	0	1	1	0	0	0	0	0	0	0
N	1752	986	766	1223	669	554	243	144	99	286	173	113

Tabella 18b - Provenienza degli alunni secondo la zona di residenza: I - II - III - IV - V superiore. Valori percentuali.

I ^A Superiore	Zona di residenza											
	ALTO			MEDIO			BASSO			TOTALE		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
Regolari normali	86.1	83.6	88.9	87.4	85.9	89.0	86.3	86.0	86.7	86.4	84.6	88.5
Regolari atipici	1.1	1.3	0.9	1.0	0.9	1.0	1.2	0.7	1.9	1.1	1.1	1.1
Ripetenti	11.2	14.0	8.0	8.5	11.3	5.4	8.8	11.1	5.9	10.2	12.8	7.1
Rientri	1.5	0.9	2.1	3.1	1.88	4.6	3.6	2.3	5.3	2.2	1.4	3.2
Non identificato	0.1	0.2	0.1	0.0	0.0	0.0	0.1	0.0	0.2	0.1	0.1	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	2895	1530	1365	833	441	392	1088	614	474	4816	2585	2231
II ^A Superiore												
Regolari normali	93.6	92.5	95.0	95.1	94.1	96.5	94.2	93.3	95.2	94.0	92.9	95.3
Regolari atipici	1.0	0.9	1.0	0.8	1.5	0.0	0.7	0.7	0.7	0.9	1.0	0.7
Ripetenti	5.1	6.6	3.5	4.0	4.5	3.5	4.7	5.6	3.7	4.8	6.0	3.5
Rientri	0.2	0.0	0.3	0.0	0.0	0.0	0.3	0.2	0.4	0.2	0.0	0.3
Non identificato	0.1	0.0	0.2	0.1	0.0	0.0	0.1	0.2	0.0	0.1	0.0	0.1
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	2721	1396	1323	772	404	367	993	536	457	4486	2336	2147
III ^A Superiore												
Regolari normali	93.0	90.7	95.7	89.4	85.2	95.2	90.3	87.5	94.5	91.8	89.1	95.4
Regolari atipici	2.6	3.6	1.4	6.6	9.4	2.6	5.3	7.0	2.7	3.8	5.3	1.8
Ripetenti	4.2	5.4	2.8	4.0	5.3	2.2	4.4	5.4	2.7	4.2	5.4	2.7
Rientri	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1
Non identificato	0.1	0.3		0.0	0.0		0.0	0.0		0.1	0.2	0.0
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	2199	1188	1010	546	318	227	733	441	292	3478	1947	1529
IV ^A Superiore												
Regolari normali	98.4	98.7	98.2	89.5	90.2	88.7	95.0	95.1	94.9	96.6	96.9	96.3
Regolari atipici	1.6	1.3	1.8	10.5	9.8	11.3	5.0	4.9	5.1	3.4	3.1	3.7
Ripetenti												
Rientri												
Non identificato												
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	1328	679	649	266	133	133	339	182	157	1933	994	939
V ^A Superiore												
Regolari normali	98.2	97.5	99.3	96.7	95.1	99.0	99.3	99.4	99.1	98.2	97.5	99.2
Regolari atipici	0.4	0.1	0.7	0.4	0.0	1.0	0.3	0.6	0.0	0.4	0.2	0.7
Ripetenti	1.2	2.1	0.0	2.9	4.9	0.0	0.3	0.0	0.9	1.3	2.1	0.1
Rientri	0.1	0.1		0.0	0.0		0.0	0.0		0.1	0.1	0.0
Non identificato	0.1	0.1		0.0	0.0		0.0	0.0		0.1	0.1	0.0
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	1223	669	553	243	144	99	286	173	113	1752	986	765

Tabella 19a - Alunni ripetenti nelle varie classi scolastiche "post-obbligo" per professione, titolo di studio del padre e zona di residenza. Valori assoluti.

Ripetenti	Professione del padre											
	Alta				Media				Bassa			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
I Superiore	42	27	15	412	209	135	74	2050	220	153	67	2429
II Superiore	33	20	13	402	83	51	32	1834	89	62	27	2132
III Superiore	24	18	6	359	53	38	15	1477	68	51	17	1589
IV Superiore	3	1	2	278	10	4	6	986	11	10	1	760
V Superiore	1	1	0	250	7	7	0	833	13	11	2	636

Ripetenti	Titolo di studio del padre											
	Alta				Media				Bassa			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
I Superiore	82	56	26	697	154	98	56	1407	284	166	78	2855
II Superiore	28	15	13	630	87	59	28	1302	99	66	33	2499
III Superiore	23	15	8	564	54	42	12	1060	77	54	23	1849
IV Superiore	5	2	3	455	10	4	6	697	12	12	0	904
V Superiore	1	1	0	415	8	7	1	587	14	14	0	747

Ripetenti	Zona di residenza											
	Alta				Media				Bassa			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
I Superiore	323	214	109	2895	71	50	21	833	96	68	28	1088
II Superiore	138	92	46	2721	31	18	13	772	47	30	17	993
III Superiore	93	64	29	2199	22	17	5	546	32	24	8	733
IV Superiore				1328				266				339
V Superiore	15	14	1	1223	7	7	0	243	286	173	113	286

Tabella 19b - Alunni ripetenti nelle varie classi scolastiche "post-obbligo" per professione, titolo di studio del padre e zona di residenza. Valori percentuali.

% Ripetenti	Professione del padre											
	Alta				Media				Bassa			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
I Superiore	10.2	11.8	8.2	412	10.2	12.6	7.6	2050	9.1	11.7	6.0	2429
II Superiore	8.2	9.3	7.0	402	4.5	5.5	3.5	1834	4.2	5.5	2.7	2132
III Superiore	7.0	9.0	3.8	359	3.6	4.9	2.2	1477	4.3	5.5	2.6	1589
IV Superiore	1.1	0.7	1.6	278	1.0	0.8	1.2	986	1.4	2.4	0.3	760
V Superiore	0.4	0.8	0.0	250	0.8	1.6	0.0	833	2.0	2.8	0.4	636

% Ripetenti	Titolo di studio del padre											
	Alta				Media				Bassa			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
I Superiore	11.8	14.9	8.1	697	10.9	13.6	8.2	684	8.5	10.8	5.9	1313
II Superiore	4.4	4.6	4.3	630	6.7	9.2	4.3	657	4.0	5.0	2.8	2499
III Superiore	4.2	5.2	2.9	564	5.1	7.5	2.4	1060	4.2	5.0	3.0	1849
IV Superiore	1.1	0.9	1.3	455	1.4	1.1	1.8	697	1.3	2.4	0.0	904
V Superiore	0.5	0.5	0.0	415	1.4	2.2	0.4	587	1.9	3.1	0.0	747

% Ripetenti	Zona di residenza											
	Alta				Media				Bassa			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
I Superiore	11.2	14.0	8.0	2895	8.5	11.3	5.4	833	8.8	11.1	5.9	1088
II Superiore	5.1	6.6	3.5	2721	4.0	4.5	3.5	772	4.7	5.6	3.7	993
III Superiore	4.2	5.4	2.8	2199	4.0	5.3	2.2	546	4.4	5.4	2.7	733
IV Superiore				1328				266				339
V Superiore	1.2	2.1	0.0	1223	2.9	4.9	0.0	243	0.3	0.0	0.9	286

Tabella 20a - Esito scolastico secondo la professione del padre, il titolo di studio del padre e il livello socio-economico della zona di residenza dell'allievo. Valori assoluti.

Respinti

	Professione del padre											
	Bassa				Media				Alta			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
III Media	139	89	50	3486	87	53	34	2488	11	4	7	420
I Superiore	313	212	101	2429	256	175	81	2047	42	32	10	412
II Superiore	201	131	70	2129	171	108	63	1832	41	22	18	400
III Superiore	133	93	40	1587	140	93	47	1474	43	29	14	357
IV Superiore	47	35	12	759	71	56	15	983	20	15	5	277
V Superiore	30	24	6	635	35	22	13	833	13	8	5	250
	Titolo di studio del padre											
III Media	157	98	59	4183	63	38	25	1563	11	6	5	678
I Superiore	366	243	123	2853	175	128	47	1406	84	57	27	697
II Superiore	233	150	83	2495	124	73	51	1299	59	41	18	629
III Superiore	141	101	40	1844	109	71	38	1060	66	42	24	563
IV Superiore	66	53	13	902	56	40	16	695	23	18	5	454
V Superiore	43	33	10	746	24	17	7	587	18	10	8	415
	Zona di residenza											
III Media	60	38	22	1715	32	17	15	1232	154	101	53	3508
I Superiore	146	101	45	1087	108	78	30	833	368	244	124	2893
II Superiore	83	44	39	992	83	53	30	770	260	176	84	2715
III Superiore	59	35	24	731	51	32	19	545	207	145	62	2191
IV Superiore	29	20	9	357	10	9	1	274	104	79	25	1381
V Superiore	8	7	1	286	14	10	4	242	55	37	18	1222

Tabella n. 20b - Esito scolastico secondo la professione, il titolo di studio del padre e il livello socio-economico della zona di residenza dell'allievo. Valori percentuali.

% Respinti

	Professione del padre											
	Bassa				Media				Alta			
	MF	M	F	N	MF	M	F	N	MF	M	F	N
III Media	4.0	5.2	2.8	3486	3.5	4.2	2.8	2488	2.6	1.8	3.6	420
I Superiore	12.9	8.7	4.2	2429	12.5	8.5	4.0	2047	10.2	7.8	2.4	412
II Superiore	9.4	6.2	3.3	2129	9.3	5.9	3.4	1832	10.3	5.5	4.5	400
III Superiore	8.4	5.9	2.5	1587	9.5	6.3	3.2	1474	12.0	8.1	3.9	357
IV Superiore	6.2	4.6	1.6	759	7.2	5.7	1.5	983	7.2	5.4	1.8	277
V Superiore	4.7	3.8	0.9	635	4.2	2.6	1.6	833	5.2	3.2	2.0	250
	Titolo di studio del padre											
III Media	3.8	4.7	2.8	4183	4.0	5.0	3.1	1563	1.6	1.7	1.5	678
I Superiore	12.8	15.8	9.4	2853	12.4	17.8	6.9	1406	12.1	15.2	8.4	697
II Superiore	9.3	11.3	7.1	2495	9.5	11.4	7.8	1299	9.4	12.5	6.0	629
III Superiore	7.6	9.4	5.2	1844	10.3	12.6	7.6	1060	11.7	14.6	8.8	563
IV Superiore	7.3	10.7	3.2	902	8.1	11.1	4.8	695	5.1	7.9	2.2	454
V Superiore	5.8	7.4	3.4	746	4.1	5.3	2.6	587	4.3	4.8	3.9	415
	Zona di residenza											
III Media	3.5	4.3	2.6	1715	2.6	2.9	2.3	1232	4.4	5.8	3.0	3508
I Superiore	13.4	9.7	3.8	1087	13.0	9.4	3.6	833	12.7	8.4	4.3	2893
II Superiore	8.4	4.4	3.9	992	10.8	6.9	3.9	771	9.6	6.5	3.1	2716
III Superiore	8.1	4.8	3.3	731	9.4	5.9	3.5	545	9.4	6.6	2.8	2192
IV Superiore	7.4	5.6	1.8	337	4.1	3.8	0.4	266	6.9	5.4	1.6	1325
V Superiore	2.8	2.4	0.3	286	5.8	4.1	1.7	242	4.5	3.0	1.5	1223

Tabella 21a - Comportamento della leva di alunni iscritti in III media nell'anno scolastico 1973/74 nel passaggio alla scuola post-obbligo. Dati globali e per professione del padre, titolo di studio del padre e zona di residenza dell'allievo. Valori assoluti.

Comportamento:

	Totali			
	M	F	MF	
Regolari serie A	352	292	644	
Regolari serie B	887	709	1596	
Regolari serie C	946	1011	1957	4197 continuano
Irregolari	1042	1212	2254	2254 Non continuano
N	3227	3224	6451	

Professione del padre

	Alta			Media			Bassa		
	MF	M	F	MF	M	F	MF	M	F
Regolari serie B	132	73	59	674	352	322	705	407	298
Regolari serie C	41	18	23	624	302	322	1158	556	602
Irregolari	44	26	18	658	337	320	1322	570	748
N	379	205	174	2273	1161	1111	3316	1608	1704

Titolo di studio del padre

Regolari serie A	278	147	131	176	89	87	158	98	60
Regolari serie B	190	102	88	529	283	246	805	454	351
Regolari serie C	78	30	48	396	173	223	1394	694	700
Irregolari	69	38	31	322	147	174	1604	726	874
N	615	317	298	1423	692	730	3961	1972	1985

Zona di residenza

Regolari serie A	456	238	218	82	54	28	89	48	41
Regolari serie B	937	524	413	267	139	128	340	188	152
Regolari serie C	962	414	548	384	185	199	529	299	230
Irregolari	981	475	504	447	176	270	696	310	385
N	3336	1651	1683	1180	554	625	1654	845	808

Tabella 21b - Comportamento della leva di alunni iscritti in III media nell'anno scolastico 1973/74 nel passaggio alla scuola post-obbligo. Dati globali e per professione del padre, titolo di studio del padre e zona di residenza dell'allievo. Valori percentuali.

	M	F	MF
Regolari serie A	10.9	9.1	10.0
Regolari serie B	27.5	22.0	24.7
Regolari serie C	29.3	31.4	30.3
Irregolari	32.3	37.6	34.9
Totale	100.0	100.0	100.0
N	3227	3224	6451

Professione del padre

	Alta			Media			Bassa		
	MF	M	F	MF	M	F	MF	M	F
Regolari serie B	34.8	35.6	33.9	29.7	30.3	29.0	21.3	25.3	17.5
Regolari serie C	10.8	8.8	13.2	27.5	26.0	29.0	34.9	34.6	35.3
Irregolari	11.6	12.7	10.3	28.8	29.0	28.8	39.9	35.4	43.9
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	379	205	174	2273	1161	1111	3316	1608	1704

Titolo di studio del padre

Regolari serie A	45.2	46.4	44.0	12.4	12.9	11.9	4.0	5.0	3.0
Regolari serie B	30.9	32.2	29.5	37.2	40.9	33.7	20.3	23.0	17.7
Regolari serie C	12.7	9.5	16.1	27.8	25.0	30.5	35.2	35.2	35.3
Irregolari	11.2	12.0	10.4	22.6	21.2	23.8	40.5	36.8	44.0
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	615	317	298	1423	692	730	3961	1972	1985

Zona di residenza

Regolari serie A	13.7	14.4	13.0	6.9	9.7	4.5	5.4	5.7	5.1
Regolari serie B	28.1	31.7	24.5	22.6	25.1	20.5	20.6	22.2	18.8
Regolari serie C	28.8	25.1	32.6	32.5	33.4	31.8	32.0	35.4	18.5
Irregolari	29.4	28.8	29.9	37.9	31.8	43.2	42.1	36.7	47.6
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	3336	1651	1683	1180	554	625	1654	845	808

Tabella 22a - Destinazione degli alunni dopo la III media in provincia di Trento secondo la professione, il titolo di studio del padre e la zona di residenza dell'alunno. Valori assoluti.

	Totali			Professione del padre								
	MF	M	F	Alta			Media			Bassa		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
Stesso Comune	1910	960	950	802	403	399	780	390	390			
Stesso Distretto	1573	873	700	579	308	271	821	463	358			
Fuori Distretto	1155	606	549	415	210	205	583	302	281			
Fuori Provincia	90	45	45	46	21	25	26	15	11			
Abbandona	2001	893	1108	570	288	292	1194	491	703			
N	6729	3377	3352	2412	1230	1182	3404	1661	1743			

	Zona di residenza			Titolo di studio del padre								
	MF	M	F	Alta			Media			Bassa		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
Stesso Comune	1682	834	848	144	73	71	55	37	18			
Stesso Distretto	647	351	296	300	157	143	575	337	238			
Fuori Distretto	308	162	146	337	171	166	417	210	207			
Fuori Provincia	25	11	14	37	20	17	25	14	11			
Abbandona	831	381	450	411	159	252	641	275	366			
N	3493	1739	1754	1229	580	649	1713	873	840			

Tabella 22b - Destinazione degli alunni dopo la III media in provincia di Trento secondo la professione, il titolo di studio del padre e la zona di residenza dell'alunno. Valori percentuali.

	Professione del padre											
	TOTALE			ALTA			MEDIA			BASSA		
	MF	M	F	MF	M	F	MF	M	F	MF	M	F
Stesso Comune	28.4	28.4	28.3	52.1	50.9	53.4	33.3	32.8	33.8	22.9	23.5	22.4
Stesso Distretto	23.4	25.9	20.9	19.2	18.6	19.9	24.0	25.0	22.9	24.1	27.9	20.6
Fuori Distretto	17.2	17.9	16.4	18.2	20.0	16.2	17.2	17.1	17.4	17.1	18.2	16.2
Fuori Provincia	1.3	1.3	1.3	2.9	1.4	4.7	1.9	1.7	2.1	0.8	0.9	0.6
Abbandona	29.7	26.4	33.1	7.5	9.1	5.8	23.6	23.4	23.8	35.1	29.6	40.3
Totale	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	6729	3377	3352	411	220	191	2412	1230	1181	3404	1661	1739

	Titolo di studio del padre								
	ALTO			MEDIO			BASSO		
	MF	M	F	MF	M	F	MF	M	F
Stesso Comune	61.2	64.2	58.1	42.6	41.7	43.5			
Stesso Distretto	16.0	14.9	17.1	22.5	25.0	20.1	18.6	18.7	18.5
Fuori Distretto	13.5	12.8	14.3	15.9	16.4	15.5	26.0	29.1	22.9
Fuori Provincia	2.1	1.2	3.1	1.7	2.0	1.3	18.6	19.6	17.7
Abbandona	7.2	6.9	7.5	17.4	14.9	19.7	1.1	1.1	1.2
Totale	100.0	100.0	100.0	100.0	100.0	100.0	35.7	31.6	39.7
N	657	335	322	1514	745	768	100.0	100.0	100.0

	Zona di residenza								
	ALTA			MEDIA			BASSA		
	MF	M	F	MF	M	F	MF	M	F
Stesso Comune	48.2	48.0	48.3	11.7	12.6	11.0			
Stesso Distretto	18.5	20.2	16.9	24.4	27.1	22.1	3.2	4.2	2.1
Fuori Distretto	8.8	9.3	8.3	27.4	29.5	25.6	33.6	38.6	28.4
Fuori Provincia	0.7	0.6	0.8	3.0	3.4	2.6	24.3	24.1	24.7
Abbandona	23.8	21.9	25.6	33.4	27.4	38.7	1.5	1.6	1.3
Totale	100.0	100.0	100.0	100.0	100.0	100.0	37.4	31.5	43.5
N	3493	1739	1752	1229	580	648	1713	873	839

Allegati

INDAGINE SULLA SITUAZIONE SCOLASTICA NEL TRENTINO

V^a Scheda alunno Scuola Media Superiore

È in corso una ricerca per meglio conoscere i problemi dei ragazzi che frequentano le scuole della provincia di Trento. Anche in passato abbiamo distribuito un questionario simile a questo; ci interessano infatti le modifiche e i cambiamenti che avvengono da un anno all'altro. Ti chiediamo perciò di riempire il questionario in ogni sua parte, assicurandoti che, come per il passato, le risposte che darai resteranno anonime.

Scuola Media Superiore _____

Tipo di scuola _____

Classe _____

Comune sede della scuola _____

1. Se potessi tornare indietro, cambieresti il tipo di scuola o, tutto sommato, sei soddisfatto di quella che stai attualmente frequentando?

Se potessi, sceglierei un altro tipo di scuola e cioè:

Tutto considerato, sono soddisfatto degli studi che sto facendo.

Mi fermerei alla licenza media inferiore.

Non saprei.

2. Quale tipo di attività lavorativa pensi di svolgere una volta terminati gli studi?

3. Pensi che la scuola che hai frequentato ti abbia fornito la preparazione necessaria per il lavoro che intendi svolgere?

Sì, senza altro

Sì, abbastanza

Non so giudicare

No

4. Dopo il diploma, pensi di continuare gli studi?

Sì No Non so

5. Se pensi di continuare a studiare, che tipo di studi intendi seguire?

Universitari

Corso di laurea di _____

Nella sede di _____

Non universitari

Diploma di _____

Nella sede di _____

6. Hai mai ripetuto qualche classe?

No, mai

Sì, una volta

Sì, due volte

Sì, tre volte

Sì, quattro volte o più

7. Hai mai ripetuto qualche classe nella tua carriera scolastica?

No Sì
1 volta Sì 2 o
+ volte

Elementare	I	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	II	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	III	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	IV	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	V	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Media	I	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	II	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	III	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Media superiore	I	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	II	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	III	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	IV	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	V	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. Quale scuola e classe frequentavi l'anno scorso?

Nome della scuola: _____

Tipo di scuola: _____

Classe: _____ Sezione: _____

Comune sede della scuola: _____

9. Dove hai frequentato la Scuola Media (inferiore)?

Scuola: _____

Comune sede della scuola: _____

INDAGINE SULLA SITUAZIONE SCOLASTICA NEL TRENTINO

Scheda alunno Scuola Media Superiore

E' in corso una ricerca per conoscere meglio i problemi dei ragazzi che frequentano le scuole della provincia di Trento. Ti chiediamo perciò di riempire il questionario in ogni sua parte, assicurandoti che le risposte che darai non verranno pubblicate.

Scuola media Superiore _____	1 2 3 4 5 6 7 8 □ □ □ □ □ □ □ □	5. La scuola che attualmente frequentavi è quella a cui desideravi veramente iscriverti oppure no?	46 47 48 □ □ □
Tipo di scuola _____		<input type="checkbox"/> 1 Sì, volevo iscrivermi a questo tipo di scuola	
Classe _____ Sezione _____		<input type="checkbox"/> 2 No, avrei preferito iscrivermi a _____	
1. Quale mezzo di trasporto usi per andare a scuola?	41 □	<input type="checkbox"/> 3 Non avevo idee precise al momento della scelta	
<input type="checkbox"/> 1 Treno		<input type="checkbox"/> 4 Ho cambiato idea (prima pensavo di fare _____)	
<input type="checkbox"/> 2 Corriera			
<input type="checkbox"/> 3 Autobus		6. Dopo il diploma, pensi di continuare gli studi?	49 □
<input type="checkbox"/> 4 Mezzi della Scuola		<input type="checkbox"/> 1 Sì <input type="checkbox"/> 2 No <input type="checkbox"/> 3 Non so	
<input type="checkbox"/> 5 Mezzo proprio			
<input type="checkbox"/> 6 nessun mezzo		7. Quale tipo di attività lavorativa pensi di svolgere una volta terminati gli studi?	50 □
2. Quanto tempo impieghi, approssimativamente, per andare a scuola?	42 □	<input type="checkbox"/> 1 Imprenditore	
<input type="checkbox"/> 1 Fino a 15 minuti		<input type="checkbox"/> 2 Dirigente	
<input type="checkbox"/> 2 da 15 a 30 minuti		<input type="checkbox"/> 3 Libero Professionista (avvocato, medico, ecc.)	
<input type="checkbox"/> 3 da 30 a 45 minuti		<input type="checkbox"/> 4 Impiegato	
<input type="checkbox"/> 4 da 45 a 60 minuti		<input type="checkbox"/> 5 Operaio o salariato	
<input type="checkbox"/> 5 più di 60 minuti		<input type="checkbox"/> 6 Lavoratore in proprio (artigiano, commerciante, ecc.)	
3. Usi fruibili di borse di studio? sussidi, ecc.?	43 □	<input type="checkbox"/> 7 Casalinga	
<input type="checkbox"/> 1 sì <input type="checkbox"/> 2 no		<input type="checkbox"/> 8 Altro _____	
4. Perché hai scelto questo tipo di scuola? (dare al massimo 2 risposte, numerate in ordine di importanza)	44 45 □ □	8. Pensi che la scuola che attualmente frequentavi ti fornirà la preparazione necessaria per il lavoro che intendi svolgere?	51 □
<input type="checkbox"/> 1 Perché tra tutti i possibili, è il tipo di studi che mi interessa di più		<input type="checkbox"/> 1 Sì, senz'altro	
<input type="checkbox"/> 2 Perché mi permette di proseguire gli studi fino alla laurea		<input type="checkbox"/> 2 Sì, abbastanza	
<input type="checkbox"/> 3 Perché è il tipo di studi che mi permette di arrivare più facilmente ad un lavoro		<input type="checkbox"/> 3 Non so giudicare	
<input type="checkbox"/> 4 Perché mi è stato consigliato da parenti o amici		<input type="checkbox"/> 4 No	
<input type="checkbox"/> 5 Perché è la scuola più vicina a casa		9. Quali sono, fra le materie che attualmente stai studiando, quelle che ti sembrano più utili ed interessanti?	52 53 □ □
<input type="checkbox"/> 6 Per altri motivi (indicali per cortesia)			
		Più utili	Più interessanti
		1° _____	1° _____
		2° _____	2° _____
		3° _____	3° _____

9a. E quali sono quelle che ti sembrano inutili o non interessanti?

Più inutili		Non interessanti	
1° _____	1° _____	2° _____	2° _____
3° _____	3° _____		

10. Quali materie ti piacerebbe venissero aggiunte nel programma, oltre a quelle esistenti?

1° _____
2° _____
3° _____

11. Ripetto alla scuola media inferiore, quella che attualmente frequenti ti sembra:

<input type="checkbox"/> 1 più facile	<input type="checkbox"/> 1 più interessante
<input type="checkbox"/> 2 più difficile	<input type="checkbox"/> 2 meno interessante
<input type="checkbox"/> 3 quasi uguale	<input type="checkbox"/> 3 quasi uguale

12. Sei ripetente?

1 sì 2 no

12a. Se sei ripetente, indica la scuola media superiore che hai frequentato l'anno scorso

Scuola _____
Indirizzo _____

12b. Se non sei ripetente, indica la scuola media inferiore e la classe che hai frequentato l'anno scorso

Scuola _____
Classe _____
Sezione _____

13. Quanto tempo dedichi allo studio (oltre a quello passato a scuola)

1 1 ora
 2 2 ore
 3 3 ore
 4 4 o più ore

14. Oltre a studiare cosa fai?

Leggo libri (ad es. _____) molto abbast. poco

Leggo fumetti (ad es. _____)

Leggo quotidiani (quale? _____)

Vado al cinema (che genere di film? _____)

Ascolto dischi o musica (che tipo di musica? _____)

Vado con gli amici o amiche

54 55

56

57 58

59

60

61

62 63 64 65 66 67

15. In quale comune abiti?

16. Professione del padre della madre

Imprenditore	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Dirigente	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Libero professionista	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Impiegato	4 <input type="checkbox"/>	4 <input type="checkbox"/>
Operaio o salariato	5 <input type="checkbox"/>	5 <input type="checkbox"/>
Lavoratore in proprio	6 <input type="checkbox"/>	6 <input type="checkbox"/>
Casalinga	7 <input type="checkbox"/>	7 <input type="checkbox"/>
Pensionato	8 <input type="checkbox"/>	8 <input type="checkbox"/>
Altro	9 <input type="checkbox"/>	9 <input type="checkbox"/>
Defunto o assente	0 <input type="checkbox"/>	0 <input type="checkbox"/>

17. Ramo di attività del padre della madre

Agricoltura	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Industria	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Commercio	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Artigianato	4 <input type="checkbox"/>	4 <input type="checkbox"/>
Pubblica Amministrazione e servizi Pubblici	5 <input type="checkbox"/>	5 <input type="checkbox"/>
Credito, assicurazioni e finanze	6 <input type="checkbox"/>	6 <input type="checkbox"/>
Altro	7 <input type="checkbox"/>	7 <input type="checkbox"/>
Nessuno	8 <input type="checkbox"/>	8 <input type="checkbox"/>

18. Titolo di studio del padre della madre

Nessun titolo	1 <input type="checkbox"/>	1 <input type="checkbox"/>
Licenza elementare	2 <input type="checkbox"/>	2 <input type="checkbox"/>
Licenza Media inferiore	3 <input type="checkbox"/>	3 <input type="checkbox"/>
Istituto professionale o elimbi	4 <input type="checkbox"/>	4 <input type="checkbox"/>
Diploma scuola media Superiore	5 <input type="checkbox"/>	5 <input type="checkbox"/>
Laurea	6 <input type="checkbox"/>	6 <input type="checkbox"/>

19. Quando sei nato?

giorno _____
mese _____
anno _____

20. Sesso

1 Maschio 2 Femmina

21. Cognome e nome

Tipo sk

68 69 70

71 72

73

76 77 78 79

80

