



Francesco Cerritelli • Diego Lanaro

Elementi di ricerca in Osteopatia e Terapie Manuali

FRANCESCO CERRITELLI DIEGO LANARO

Elementi di ricerca in osteopatia e terapie manuali



Francesco Cerritelli, Diego Lanaro
ELEMENTI DI RICERCA IN OSTEOPATIA E TERAPIE MANUALI
Copyright © 2018, EdiSES s.r.l. – Napoli

9 8 7 6 5 4 3 2 1 0
2022 2021 2020 2019 2018

Le cifre sulla destra indicano il numero e l'anno dell'ultima ristampa effettuata

A norma di legge è vietata la riproduzione, anche parziale, del presente volume o di parte di esso con qualsiasi mezzo.

L'Editore

L'editore ha effettuato quanto in suo potere per richiedere il permesso di riproduzione del materiale di cui non è titolare del copyright e resta comunque a disposizione di tutti gli eventuali aventi diritto.

AVVERTENZA: la medicina è una scienza in costante progresso. La ricerca assidua e l'esperienza clinica portano necessariamente a instancabili cambiamenti e ad approcci terapeutici e assistenziali sempre più moderni e sicuri. L'autore ha posto la massima attenzione per garantire l'accuratezza di tutte le informazioni contenute in quest'opera; nonostante ciò non può garantire che le informazioni contenute siano esaustive in ogni parte. Gli utilizzatori dovranno controllare attentamente e verificare le informazioni scientifiche contenute e assicurarsi che non siano intervenute nel frattempo nuove indicazioni, controindicazioni o modifiche alle procedure descritte nell'opera.

Pur avendo garantito la massima cura e il massimo impegno in ciascuna delle fasi di pubblicazione, l'autore, l'editore, i curatori e i distributori declinano ogni responsabilità circa errori o omissioni o per qualsiasi conseguenza derivante dall'applicazione delle informazioni di questa opera e non danno nessuna garanzia, esplicita o implicita, rispetto al contenuto della pubblicazione.

Impaginazione:

Vcolore di Francesco Omaggio - Pordenone

Stampato presso:

Vulcanica s.r.l. Nola, Napoli

per conto della:

EdiSES S.r.l – Piazza Dante, 89 – Napoli

www.edises.it info@edises.it

ISBN 978 88 3319 011 2

AUTORI

ANDREA BERGNA, PT D.O.

SOMA Istituto Osteopatia Milano.

Fisioterapista dal 1993 e Osteopata dal 1997, specializzato nell'ambito ortopedico, traumatologico e sportivo; appassionato all'insegnamento e alla ricerca della pratica osteopatica. Socio fondatore di SOMA - Istituto Osteopatia Milano dal 2000 e coautore del *Core Competence dell'Osteopata* redatto dal Registro degli Osteopati d'Italia.

FRANCESCA BOVIS, PhD

Dipartimento di Scienze della Salute, Università degli studi di Genova.

Lavora come Assegnista di ricerca presso la sezione di Biostatistica del Dipartimento di Scienze della Salute (DISSAL) dell'Università di Genova. Il suo ambito di ricerca si è focalizzato sull'analisi statistica di dati medici e in particolare sui problemi metodologici negli studi di sclerosi multipla.

FRANCESCO CERRITELLI, PhD MPH D.O. MS

Fondazione C.O.ME. Collaboration ONLUS; Università Chieti-Pescara.

DO, PhD in neuroscienze all'Università di Chieti-Pescara, è presidente di C.O.ME. Collaboration Onlus. Specializzato nella ricerca osteopatica, si è focalizzato sui meccanismi sottesi all'efficacia del trattamento osteopatico e sugli aspetti interdisciplinari dell'EBM.

MARCO CHIERA, Naturopata - Master II livello in PNEI e Scienza della Cura Integrata,

Fondazione C.O.ME. Collaboration ONLUS; Membro Gruppo Coordinamento Dis-Co SIPNEI.

Laureato in Filosofia e poi in Scienze Cognitive e Processi Decisionali, lavora come naturopata dal 2014. Co-autore di due libri (*La PNEI e il sistema miofasciale* e *La PNEI e le discipline corporee*), che intendono fornire una solida base scientifica alle pratiche terapeutiche che coinvolgono il corpo.

MASSIMILIANO GRASSI, M.Sc.

Dipartimento di Neuroscienze Cliniche, Casa di Cura Villa San Benedetto Menni, Suore Ospedaliere, FoRiPsi; Albese con Cassano, Como. Istituto Superiore di Osteopatia, Milano.

Ricercatore, statistico, psicoterapeuta ad orientamento cognitivo e docente di statistica e machine learning. Autore di oltre 15 pubblicazioni su riviste scientifiche internazionali, si interessa attualmente dello sviluppo di algoritmi di intelligenza artificiale per la medicina di precisione.

STEFANO GUALDI, Informatico

Internet Express.

Dopo una formazione culturale di carattere tecnico, da circa 25 anni svolge attività di consulenza e sviluppo software. Molto attento al mondo *open source*, sviluppa soluzioni per i suoi clienti con le più recenti tecnologie disponibili sul mercato.

ALESSIO IACOPINI, MSc Ost, D.O., BSc MS

Fondazione C.O.ME. Collaboration ONLUS; EDUCAM-Dipartimento Internazionali.

Laureato in Scienze Motorie (Master of Sciences in Osteopatia) ed Osteopata D.O., insegna manipolazione fasciale e razionale osteopatico in diversi istituti di formazione italiani ed europei. Si occupa principalmente di ricerca qualitativa nel campo delle terapie manuali.

ELEONORA LACORTE, Ricercatore

Centro nazionale per la prevenzione delle malattie e la promozione della salute, Istituto Superiore di Sanità.

Ricercatore presso il Reparto Promozione e valutazione delle politiche di prevenzione delle malattie croniche del Centro nazionale per la prevenzione delle malattie e la promozione della salute, specializzata in metodologia per la produzione e valutazione di linee guida, revisioni sistematiche e metanalisi.

DIEGO LANARO, PhD D.O.

Fondazione C.O.ME. Collaboration ONLUS

Osteopata D.O., Biologo Ph.D., docente IEMO e Università di Siena, Genova e Torino (formazione post-laurea), docente Osteopatia per Bambini, Membro Direttivo SIPNEI Liguria, Coordinatore Gruppo Osteopati Dis-Co SIPNEI, Membro Direttivo C.O.ME. Collaboration Onlus, Istruttore di Taiji Nei Dan School ASI-Coni.

CHRISTIAN LUNGI, ND D.O.

Fondazione C.O.ME. Collaboration ONLUS.

Alla ricerca... di una mente che danza in un corpo che pensa per dare il mio contributo nella famiglia C.O.ME., nei gruppi Core Competence ROI e Curriculum AISO, nelle parole e nei numeri di docenze, congressi, libri, pubblicazioni e nella cura osteopatica.

MARCO PETRACCA, Osteopata, Fisioterapista, Specialista in Scienze delle Professioni Sanitarie della Riabilitazione

Ospedale San Pietro FBF, Roma; Scuola di Osteopatia C.E.R.D.O., Roma; Università degli studi di Roma "Sapienza"; Registro degli Osteopati d'Italia - R.O.I.

Nato a Roma, dal 2000 è Fisioterapista e nel 2009 diventa Osteopata presso il C.E.R.D.O., dove è responsabile del Dipartimento Ricerca. Autore di diverse pubblicazioni e docente presso la "Sapienza", lavora nell'U.O.C. di Pediatria dell'Ospedale San Pietro ed è Direttore del centro Osteobimbo.

FRANCESCO RASO, PhD

Azienda Sanitaria Locale N.3 Genovese. S.C. Patologia Clinica.

Biologo Molecolare, specializzato in Patologia clinica. Ricercatore in neuroscienze e docente IEMO. Dirigente Sanitario presso la ASL3 di Genova.

SILVIA CLARA TUSCANO, Traduttrice

Fondazione C.O.ME. Collaboration ONLUS.

Traduttrice editoriale e mediatrice linguistica. Docente IEMO, Istituto Europeo per la Medicina Osteopatica. Membro C.O.ME. Collaboration Fondazione Onlus e SIPNEI Liguria.

NICOLA VANACORE, Ricercatore, PhD MD



Centro Nazionale della Prevenzione delle Malattie e Promozione della Salute, Istituto Superiore di Sanità.

Ricercatore presso il Centro Nazionale di Prevenzione delle Malattie e Promozione della Salute, Istituto Superiore di Sanità. Si occupa dell'applicazione dei metodi epidemiologici alla neurologia. È stato responsabile del Sistema Nazionale delle Linee Guida – ISS. Docente di Metodologia della Ricerca presso l'Università degli Studi di Roma “Sapienza”.

INDICE GENERALE

CAPITOLO 1


La medicina basata sulle evidenze in osteopatia

1.1	Introduzione all'evidence-based medicine	1
1.1.1	Perché EBM	2
1.1.2	Teoria dell'EBM	3
1.1.3	Evidence-based practice	4
1.1.4	Osteopatia ed EBM	12
1.2	Chi fa EBM	13
1.2.1	Le revisioni Cochrane	14
1.2.2	Ricerca osteopatica in USA, Europa e resto del mondo	16
1.3	Evoluzione dell'EBM	18
1.3.1	Evidence-informed practice	18
1.3.1.1	Evidence-informed osteopathy	19
	Conclusioni	20
	Applicazioni pratiche	22
	BIBLIOGRAFIA	22

CAPITOLO 2

Epidemiologia



	Introduzione	25
2.1	Epidemiologia ed evidence-based medicine	26
2.1.1	Principi di base	27
2.1.2	Parametri e metodi	28
2.1.3	Prevalenza e incidenza	28
2.1.3.1	Tasso di prevalenza	29
2.1.3.2	Tasso di incidenza	29
2.2	Studi osservazionali descrittivi	30
2.3	Studi osservazionali analitici: ecologici, trasversali, di coorte e caso-controllo	31

2.3.1	Studi ecologici	31
2.3.2	Studi trasversali (Cross Sectional Design)	32
2.3.3	Studi di coorte (Cohort Design)	33
2.3.4	Studi caso-controllo (Case-control Design)	35
2.4	Studi randomizzati controllati (Randomized Controlled Trials - RCTs)	37
2.5	Bias, confondimento, fattori di rischio ed esiti	39
2.5.1	Bias	39
2.5.2	Confondimento	40
2.5.3	Fattori di rischio	41
2.5.3.1	Studio retrospettivo: l'Odds Ratio	42
2.5.3.2	Studio prospettivo: il Rischio Relativo	43
2.5.3.3	Interpretazione dell'Odds Ratio e del Rischio Relativo	43
2.5.3.4	Differenza del rischio	44
2.5.4	Esiti	44
2.6	Misurabilità nel trattamento manipolativo osteopatico	45
2.7	Misurazioni degli esiti nel trattamento manipolativo osteopatico	49
	Conclusioni	52
	BIBLIOGRAFIA	53

CAPITOLO 3

Sistemi informatici



	Introduzione	57
3.1	Il software open source	59
3.1.1	Quali software open source?	60
3.2	La raccolta dei dati	60
3.2.1	Gestione dei dati raccolti	63
3.3	Uso di un database relazionale	64
3.3.1	Interazione con il database	64
3.3.2	Importazione dati	72
3.4	Analisi dei dati	72
3.5	Il sistema R	74
3.5.1	Il sistema dei pacchetti	74
3.5.2	Download ed installazione	76
3.5.3	Interazione	77
3.5.4	Nozioni base del linguaggio R	82
3.5.5	Manipolazione dei dati	91
3.5.6	La creazione di funzioni	104



3.5.7	Salvataggio e scrittura su file	105
3.5.8	Caricamento dei dati da fonti esterne	106
Conclusioni		109
	Applicazioni pratiche	110
	BIBLIOGRAFIA	111

CAPITOLO 4

Statistica di base

Introduzione		113
4.1	Analisi descrittive	114
4.1.1	Variabili statistiche	114
4.1.2	Indici descrittivi	115
4.1.2.1	Le frequenze	115
4.1.2.2	Gli indici di posizione: media, moda e mediana	117
4.1.2.3	Gli indici di dispersione: i percentili e la varianza	120
4.2	Rappresentazioni grafiche	122
4.3	Test inferenziali	127
4.3.1	Variabili casuali campionarie	127
4.3.2	I test di verifica d'ipotesi	130
4.3.3	I test di Normalità	134
4.3.4	I test di omogeneità delle varianze	139
4.4	Gli intervalli di confidenza	140
4.5	I test di significatività	142
4.6	Confrontare più gruppi fra loro	150
4.6.1	Test di verifica d'ipotesi fra gruppi	150
4.6.1.1	Test t a misure ripetute: quando due gruppi sono lo stesso gruppo	155
4.6.2	Le varianze fra e intra-gruppi	157
4.6.3	L'analisi della varianza (ANOVA)	158
4.6.3.1	L'analisi della varianza (ANOVA) ad una via	160
4.6.3.2	L'analisi della varianza (ANOVA) a due vie	161
4.6.3.3	L'analisi della varianza (ANOVA) a misure ripetute	167
4.6.3.4	La mixed ANOVA: gli studi sperimentali	172
4.7	La correlazione fra variabili	183
4.7.1	L'indice di correlazione di Pearson e l'indice di connessione χ^2	184
4.7.2	I test non parametrici di Spearman e Kendall per la correlazione	187
4.7.3	La concordanza fra due misurazioni (o operatori o test)	192
4.7.4	La concordanza fra più misurazioni (o esaminatori)	195



4.7.5	Il coefficiente di correlazione intra-classe o ICC	198
4.8	Calcolo della numerosità campionaria	207
4.8.1	Numerosità campionaria per stime di valori	207
4.8.2	Numerosità campionaria per la rilevazione di differenze	208
4.9	Sensibilità e specificità	213
	Conclusioni	218
	Applicazioni pratiche	221
	BIBLIOGRAFIA	266
CAPITOLO 5		
Tipi di studio		
	Introduzione	269
5.1	Questioni Metodologiche	270
5.1.1	La causalità	270
5.1.2	Il metodo ipotetico-deduttivo	271
5.1.3	Il metodo qualitativo	272
5.1.4	Il metodo misto	274
5.1.5	Tradurre la ricerca nella pratica	275
	5.1.5.1 Lo strumento RE-AIM	276
	5.1.5.2 Lo strumento PRECIS-2	277
5.1.6	La verifica dell'efficacy e dell'effectiveness	280
5.1.7	La generalizzabilità dei risultati	282
5.1.8	Gli errori	284
5.1.9	Il mascheramento	285
5.1.10	L'effetto placebo e l'effetto nocebo	286
5.2	I tipi di studio	287
5.2.1	Studi osservazionali	289
	5.2.1.1 Studi di coorte	290
	5.2.1.2 Studi caso-controllo	290
	5.2.1.3 Studio cross-over	291
	5.2.1.4 Studio trasversale (cross-sectional)	291
5.2.2	Studi sperimentali	292
	5.2.2.1 Studio clinico non controllato e/o non randomizzato	292
	5.2.2.2 Studio clinico controllato e randomizzato	292
5.2.3	Studi sull'accuratezza diagnostica	294
5.2.4	Gli studi N-of-1	295
5.3	Gli studi nella medicina osteopatica	296
5.4	Applicazione – Come valutare criticamente uno studio clinico	300
	Conclusioni	304

	Applicazioni pratiche: la conduzione di uno studio	305
	BIBLIOGRAFIA	310

CAPITOLO 6

La ricerca qualitativa in ambito osteopatico



6.1	Dalla ricerca quantitativa alla ricerca qualitativa	319
6.1.1	La questione del paradigma	320
6.1.2	Positivismo e metodi quantitativi	321
6.1.3	Interpretativismo e metodi qualitativi	321
6.1.4	Tipi di dati	322
6.1.5	Relazione tra dati qualitativi e analisi	323
6.2	Disegno di uno studio qualitativo	324
6.2.1	Ruolo del ricercatore nella definizione del quadro concettuale	324
6.2.2	Revisione della letteratura	325
6.2.3	Definizione del quesito di ricerca	325
6.2.4	Selezione del campione	326
6.2.5	Metodi di raccolta dei dati	327
6.2.5.1	Intervista singola	327
6.2.5.2	Intervista di gruppo e focus group	327
6.2.6	Metodi di analisi dei dati	327
6.2.7	I diversi approcci in ambito qualitativo	328
6.2.7.1	Grounded theory	328
6.2.7.2	Etnografia	330
6.2.7.3	Fenomenologia	330
6.2.7.4	Studio di caso	330
6.2.7.5	Metodi di sviluppo di consenso	331
6.2.8	Criteri di valutazione di uno studio qualitativo	331
6.2.9	Selezione della tipologia di studio qualitativo	332
6.3	Applicazioni della ricerca qualitativa in ambito osteopatico	334
6.4	Grounded theory. Descrizione generale ed applicazioni in ambito osteopatico	334
6.4.1	Cenni generali	334
6.4.2	Revisione della letteratura nell'ambito osteopatico e delle terapie manuali	336
6.4.3	Struttura di uno studio GT	336
6.4.3.1	Raccolta (campionamento intenzionale e teoretico) ed analisi simultanea dei dati, per la costruzione di codici e categorie	337
6.4.3.2	Utilizzo di metodi di analisi comparative e sviluppo di teorie innovative	337
6.4.3.3	Memo-writing	338
6.4.3.4	Modalità di analisi dei dati	338
6.4.3.5	Core Category e produzione del rapporto di ricerca	339

6.4.4	Considerazioni finali sugli studi GT	339
6.5	Delphi panel. Descrizione generale ed applicazioni in ambito osteopatico	340
6.5.1	Cenni generali	340
6.5.2	Revisione della letteratura nell'ambito osteopatico e delle terapie manuali	340
6.5.3	Struttura di uno studio DP	341
6.5.3.1	Svolgimento dei rounds	342
6.5.3.2	Punti di forza e debolezza di uno studio DP	342
6.5.4	Considerazioni finali sugli studi DP	342
6.6	Focus group. Descrizione generale ed applicazioni in ambito osteopatico	343
6.6.1	Cenni generali	343
6.6.2	Revisione della letteratura nell'ambito osteopatico e delle terapie manuali	343
6.6.3	Struttura di uno studio FG	344
6.6.3.1	Fasi di un FG	344
6.6.3.2	Il gruppo: numerosità e composizione	344
6.6.3.3	Conduzione di un FG	344
6.6.3.4	La conduzione del gruppo	345
6.6.3.5	Registrazione e analisi dei dati	345
6.6.3.6	Il report finale	345
6.6.4	Considerazioni finali sugli studi FG	345
	Conclusioni	346
	Applicazioni pratiche	347
	BIBLIOGRAFIA	348

CAPITOLO 7



La ricerca bibliografica

7.1	Significato della ricerca bibliografica	353
7.2	La letteratura scientifica	354
7.2.1	Forme di letteratura biomedica	355
7.2.2	Internet	355
7.3	Processo di pubblicazione di un articolo scientifico	356
7.3.1	Il sistema peer-review	356
7.3.2	Indicatori bibliometrici	357
7.3.3	L'articolo scientifico	359
7.4	Razionale: perché fare un'accurata ricerca bibliografica	361
7.4.1	Strategie	361
7.4.2	Motori di ricerca	363
7.4.3	I database	363

7.4.4	Parole chiave	364
7.4.5	Operatori booleani	365
7.4.6	Come limitare una ricerca su PubMed	367
7.5	Un'efficace ricerca bibliografica	367
7.6	Valutazione critica delle informazioni	369
	Conclusioni	370
	Applicazioni pratiche	371
	BIBLIOGRAFIA	374



CAPITOLO 8

Revisioni sistematiche e meta-analisi

	Introduzione	377
8.1	Cosa vuol dire revisione sistematica	379
8.1.1	Generalità	379
8.1.2	Realizzazione di una revisione sistematica	383
8.1.2.1	Formulazione del quesito e sviluppo dei criteri di inclusione	384
8.1.2.2	Ricerca e selezione degli studi	385
8.1.2.3	Valutazione dei rischi da distorsione negli studi selezionati	387
8.1.2.4	Estrazione ed analisi dei dati con eventuale meta-analisi	389
8.1.2.5	Gestione dei rischi legati al publication bias	390
8.1.2.6	Presentazione dei risultati	391
8.1.2.7	Interpretazione dei risultati	392
8.2	Principi e metodi della meta-analisi	392
8.2.1	Aspetti generali	392
8.2.2	Esecuzione della meta-analisi	393
8.2.3	Dimensione dell'effetto e precisione	394
8.2.4	Eterogeneità	395
8.2.5	Modello a effetti fissi contrapposto a modello a effetti casuali	396
8.2.6	Analisi di sensibilità e di sottogruppo	397
8.2.7	Meta-analisi cumulative	397
8.3	Valutare una revisione sistematica	398
8.4	Applicazione: meta-analisi in OMT e confronto con altre meta-analisi non OMT	402
	Conclusioni	405
	Applicazioni pratiche	408
	BIBLIOGRAFIA	408


CAPITOLO 9

Statistica avanzata

Introduzione	413
9.1 Analisi di regressione lineare semplice	414
9.2 Regressione lineare multipla	417
9.3 Regressione logistica	418
9.4 Le varianti dell'ANOVA	421
9.4.1 Analisi della covarianza (ANCOVA)	421
9.4.2 Analisi multivariata della varianza (MANOVA)	425
9.4.3 Analisi multivariata della covarianza (MANCOVA)	427
9.5 Analisi di sopravvivenza	428
9.5.1 Il metodo delle tavole di sopravvivenza	429
9.5.2 Il metodo di Kaplan-Meier	430
9.5.3 Modello di Cox	431
Conclusioni	432
 Applicazioni pratiche	433
 BIBLIOGRAFIA	441



CAPITOLO 10

Ricerca sui sistemi sanitari

Introduzione	443
10.1 Il problema della valutazione dei servizi	444
10.2 Popolazione di riferimento	444
10.3 Le possibili distorsioni	447
10.4 Fonti di variabilità	450
10.5 Ruolo delle linee guida	452
10.6 Metodi e strumenti di analisi	454
10.6.1 Health economics	454
10.6.1.1 Health economics: basi e criteri	455
10.6.2 I costi: concetti base	456
10.6.3 Le valutazioni economiche: misurare i costi	457
10.6.4 Misurazione dei costi nei sistemi sanitari: implicazioni pratiche	458
10.7 La valutazione dei servizi: quale ruolo per l'osteopatia	459
Conclusioni	461
 BIBLIOGRAFIA	462

CAPITOLO 11

I sistemi di misurazione delle performance

Introduzione	467
11.1 La performance: basi concettuali e scopi	468
11.2 Disease management	470
11.3 Sistemi di indicatori	472
11.4 Strumenti per il benchmarking dei servizi sanitari	477
11.4.1 I Benchmarking Controlled Trials: la nuova frontiera degli studi osservazionali	478
Conclusioni	480
 Applicazioni pratiche	482
 BIBLIOGRAFIA	482
Appendici	485
Indice analitico	515

PREFAZIONE

È con vero piacere che mi accingo a scrivere la prefazione di questo libro sulla ricerca in osteopatia a cura di Francesco Cerritelli e Diego Lanaro. Le ragioni sono molteplici. Come ricercatori dell'Istituto Superiore di Sanità abbiamo la missione di sostenere il raggiungimento di elevati standard di qualità di tutti i professionisti sanitari che operano per il bene primario dei cittadini, quello della salute. L'osteopatia italiana negli ultimi dieci anni ha compiuto uno sforzo notevole per l'inserimento nel proprio paradigma culturale della logica della medicina basata sulle evidenze (EBM). Questo ha significato semplicemente l'accettazione della sfida di produrre evidenze che documentino l'efficacia di un trattamento osteopatico. Accettare nella propria pratica clinica il meccanismo dell'onere della prova comporta un cambiamento radicale nella relazione terapeutica con un paziente. Condividere gli obiettivi terapeutici e le incertezze rende oggi moderno un professionista sanitario. Il paradigma dell'EBM in osteopatia non si contrappone alla storia della disciplina, a Still e a tutti coloro che hanno diffuso nel mondo quei principi. Anzi, al contrario, l'EBM completa e rilancia questa professione verso ambiti terapeutici nuovi (vedi le tecniche viscerali ad esempio) fornendo linfa vitale nell'interazione con gli altri professionisti sanitari e favorendo quell'approccio multidisciplinare per la diagnosi e la gestione delle patologie complesse. La pratica dell'EBM richiede oggi ad un professionista sanitario di saper leggere criticamente un articolo scientifico, ponderando la contaminazione tra ricerca quantitativa e qualitativa, e di trasferire con senso clinico quelle evidenze nella pratica quotidiana considerando quali sono i risultati desiderati da un paziente. Tutto ciò non è altro che una straordinaria fusione tra arte e scienza. La medicina non è una scienza esatta. Nessuno vuole ridurre la complessità di una persona e della sua condizione di salute o di malattia ad un valore numerico di una scala o a un valore statistico. Bisogna ricordare che l'EBM nasce come metodo innovativo per formare i professionisti sanitari. Ha rappresentato e rappresenta una vera rivoluzione culturale rispetto alle pratiche mediche del secolo scorso.

Questo processo culturale nell'osteopatia è stato scandito da alcuni accadimenti cruciali quali il convegno OMS a Milano del 2007, che ha definito i criteri per la formazione dell'osteopata, enfatizzando l'importanza della metodologia della ricerca, e l'inserimento nel Sistema Nazionale Americano delle Linee Guida (LG) (National Guideline Clearinghouse) dell'Associazione Americana degli Osteopati sul "low back pain" non specifico, sia acuto che cronico. In Italia ormai decine e decine di osteopati pubblicano le loro ricerche su riviste peer-review censite su Medline e disponibili su PubMed. Ed ecco che finalmente viene pubblicata sulla Gazzetta Ufficiale del 31 gennaio 2018 la Legge 3 che prevede all'articolo 7 comma 1 l'istituzione in Italia della professione sanitaria dell'osteopata. In questo contesto va quindi segnalata la necessità anche per l'osteopata di seguire le LG nella propria pratica clinica secondo quanto previsto dalla Legge 24 del 2017 sulla responsabilità professionale degli esercenti delle professioni sanitarie. I processi culturali hanno però i loro tempi ed hanno sicuramente bisogno di strumenti adeguati per poter diffondersi capillarmente. Ritengo che

questo libro abbia questo specifico compito. Nel panorama editoriale italiano vi era una grande necessità di un libro dedicato alla ricerca in osteopatia. Il piano dell'opera è articolato in 11 capitoli che spaziano dalla statistica all'epidemiologia, dalla ricerca qualitativa alla revisione sistematica e meta-analisi per completare il percorso con un focus sui sistemi sanitari e la misurazione della loro performance. Sono convinto che questo libro avrà tutto il successo che merita tra gli studenti in osteopatia e tra tutti i professionisti sanitari, osteopati e non, che potranno apprezzare gli sforzi degli autori di condurre il paradigma dell'EBM nell'alveo della pratica osteopatica.

NICOLA VANACORE

**Centro Nazionale di Prevenzione delle Malattie
e Promozione della Salute.
Istituto Superiore di Sanità.**

9

Francesca Bovis

Statistica avanzata

O B I E T T I V I

Da questo capitolo il lettore dovrebbe:

- Capire come descrivere la relazione tra una o più variabili indipendenti e una o più variabili dipendenti, tenuto conto dei possibili fattori confondenti
- Comprendere come i fattori di rischio sono associati alla presenza o all'assenza di una malattia
- Studiare la distribuzione del tempo di comparsa di un evento

Introduzione

L'obiettivo generale delle tecniche di analisi di regressione multipla è approfondire la relazione tra più variabili indipendenti (o predittori) e una variabile dipendente.

Lo scopo principale di questa sezione è fornire un'introduzione semplice e il più possibile esaustiva che permetta a chi non ha grande esperienza con i metodi di regressione di capirne e interpretarne i risultati.

Cominceremo considerando le tecniche di regressione lineare in modo da presentare i concetti di base e poi descriveremo la regressione logistica e presenteremo un esempio clinico illustrando ogni tecnica di regressione.

Tratteremo poi le varianti più complesse dell'ANOVA, ovvero: l'analisi della covarianza (*ANCOVA*), utile per considerare nel modello dell'ANOVA variabili che possono confondere o mediare l'effetto della variabile indipendente su quella dipendente; le versioni multivariate (*MANOVA* e *MANCOVA*), strumenti che permettono di valutare le influenze delle variabili indipendenti su più variabili dipendenti.

9.1 Analisi di regressione lineare semplice

Il termine “regressione” ha la sua origine nel XIX secolo quando lo statistico Francis Galton pubblicò un articolo in cui dimostrava come “ogni caratteristica di un individuo è ereditata dalla prole, ma in media ad un livello minore”. Nell’articolo Galton osservò che i figli di genitori di bassa statura tendevano ad essere più alti dei propri genitori e i figli di genitori alti erano anch’essi alti, ma in media meno del genitore. Tale fenomeno fu chiamato regressione e da allora tale termine è rimasto per descrivere quelle *tecniche statistiche che analizzano la relazione tra due o più variabili*.

Esistono differenti tipologie di regressione come mostrato in **Tabella 9.1**.

Tabella 9.1 Tipi di regressione e loro uso. Le variabili risposta sono solitamente le variabili dipendenti, mentre le variabili predittive sono quelle indipendenti. Si usano i termini “risposta” e “predittive” perché più corretti statisticamente: le variabili indipendenti possono non essere veramente tali, così come le variabili dipendenti

Tipo di regressione	Uso tipico
Lineare semplice	Predire una variabile risposta numerica da una variabile predittiva numerica
Polinomiale	Predire una variabile risposta numerica da una variabile predittiva numerica, quando la relazione segue un modello polinomiale n-esimo
Multipla lineare	Predire una variabile risposta numerica da una o più variabili predittive
Multivariata	Predire una o più variabili risposta numeriche da una o più variabili predittive
Logistica	Predire una variabile risposta categorica da una o più variabili predittive
Poisson	Predire una variabile risposta che funge da contatore di dati da una o più variabili predittive
Cox	Predire il tempo di un evento (es.: morte, ricaduta) da più variabili predittive
Time-series	Modellare i dati di una serie temporale (o storica) con gli errori di correlazione
Non lineare	Predire una variabile risposta numerica da una o più variabili predittive, quando la forma del modello non è lineare
Non parametrica	Predire una variabile risposta numerica da una o più variabili predittive, quando la forma del modello (la distribuzione) è derivata dai dati e non è specificata a priori
Robusta	Predire una variabile risposta numerica da una o più variabili predittive, usando un approccio resistente all’effetto delle osservazioni confondenti

In questo capitolo ci focalizzeremo sulla regressione semplice e multivariata (che rientrano anche nella grande classe delle regressioni OLS – *Ordinary Least Squares* – ad oggi molto utilizzate) e poi tratteremo la regressione logistica e le analisi di sopravvivenza, che include l’analisi Cox.

Nella regressione lineare il termine “semplice” si riferisce al fatto che viene studiata la relazione tra due sole variabili (questa tecnica è detta anche bivariata), mentre il termine “lineare” indica che la relazione può essere descritta da una linea retta.

L'analisi di regressione lineare è dunque una tecnica che permette di analizzare la relazione lineare tra una variabile dipendente (o variabile risposta) e una o più variabili indipendenti (o predittori). Lo studio di questa relazione può avere un duplice scopo:

- esplicativo: comprendere e ponderare gli effetti delle variabili indipendenti sulla variabile dipendente in funzione di un determinato modello teorico;
- predittivo: individuare una combinazione lineare di variabili indipendenti per predire in modo ottimale il valore assunto dalla variabile dipendente.

Per studiare la relazione tra due variabili è utile lo *scatter plot* (il diagramma di dispersione) in cui si riportano i valori della variabile esplicativa x sull'asse delle ascisse e i valori della variabile dipendente y sull'asse delle ordinate (**Figura 9.1**).

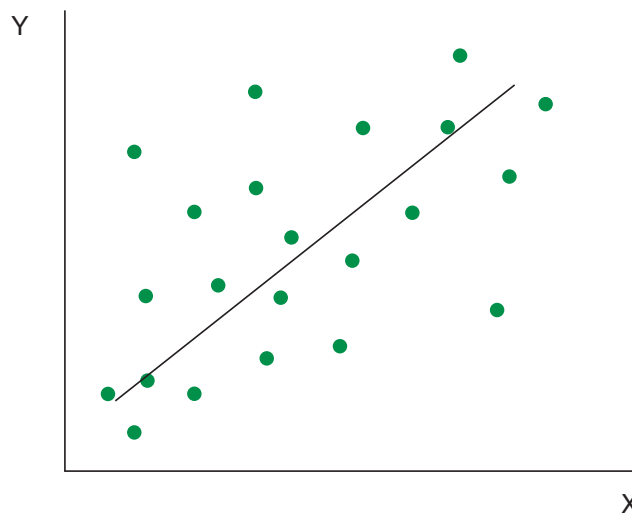


Figura 9.1 Esempio di scatter plot. Per approfondimenti, rimandiamo al capitolo 4 e alla Figura 4.1.

L'equazione della retta di regressione è data da:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

dove β_0 e β_1 sono i coefficienti di regressione della retta, o parametri, ed ε è l'errore che si commette nella spiegazione della variabile y tramite una funzione lineare di x .

β_0 rappresenta l'intercetta, ovvero il punto in cui la retta incrocia l'asse delle ordinate e β_1 rappresenta l'inclinazione della retta di regressione di y su x e indica di quante unità cambia y per una variazione unitaria che si verifica nella x .

Il segno di β_1 indica se la relazione lineare è positiva o negativa.

L'analisi della regressione lineare semplice individua la retta che consente di prevedere al meglio i punteggi nella variabile dipendente a partire da quelli della variabile indipendente. Si tratta di individuare la retta che "interpola" meglio la nuvola di punti (lo *scatter plot*) definita dalla distribuzione congiunta delle due variabili.

Le variabili, per poter essere inserite in un modello di regressione lineare, devono rispettare determinate condizioni. Primo, la variabile dipendente deve essere continua, ossia può assumere tutti gli infiniti valori appartenenti ad un intervallo di numeri reali e deve essere misurata su una scala numerica (l'età o la durata della malattia sono esempi di variabili continue).

Secondo, la relazione tra le variabili è una dipendenza funzionale. Ciò significa che la variabile dipendente (y), che di solito viene rappresentata graficamente sull'asse delle ordina-

te, è determinata da, o è funzione di, una variabile indipendente (x) rappresentata sull'asse delle ascisse.

Terzo, i valori delle x variabili indipendenti sono misurati con errore trascurabile, mentre i valori delle variabili y vengono considerati un campione casuale di una popolazione distribuita normalmente. Inoltre la deviazione standard dei valori y è uguale per ogni valore di x .

Come abbiamo scritto in precedenza, per l'analisi di regressione semplice si tratta di individuare la retta che meglio interpola la nuvola di punti, ma per un insieme di punti possono passare infinite rette di regressione. Come scegliere la migliore?

Il *metodo dei minimi quadrati* permette di scegliere la retta che minimizza la somma degli scarti dalla retta di regressione (**Figura 9.2**).

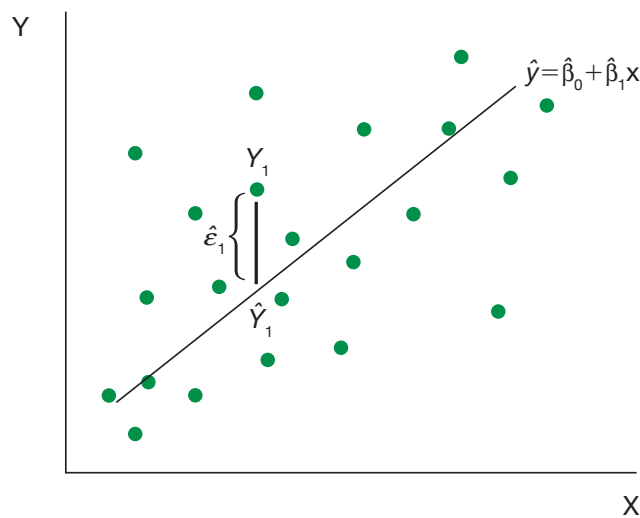


Figura 9.2 Esempio di linea di regressione, accanto alla quale è mostrato lo scarto di un punto dalla linea stessa.

In questo capitolo non tratteremo come effettuare i calcoli matematici necessari ad adattare una retta ad un insieme di dati, ma ci concentreremo piuttosto su come interpretare l'output dei risultati sulla relazione tra le variabili dipendenti e indipendenti ottenuti con l'utilizzo di software statistici dedicati.

Oggi tutti i software statistici disponibili effettuano l'analisi di regressione e l'output dei risultati prodotti contiene in generale quattro classi di informazioni:

1. l'informazione che definisce la linea retta che meglio descrive o interpola i valori della variabile dipendente per ogni valore della variabile indipendente;
2. la misura di quanto strettamente le due variabili sono correlate;
3. la probabilità che i dati, pur mostrando una relazione, siano stati tratti da una popolazione dove di fatto tale relazione non esiste;
4. la quarta classe di informazioni contiene le stime della variabilità della retta di regressione e dei suoi coefficienti (gli errori standard), e alcuni valori utilizzati nel calcolo della significatività statistica.

Consideriamo nel dettaglio le prime tre classi di informazioni e l'output fornito dal software statistico gratuito R (usando un dataset di esempio dal nome `alligator`).

```
> summary(alli.mod1)
```

```
call:
```

```
lm(formula = lnWeight ~ lnLength, data = alligator)

Residuals:
Min 1Q      Median 3Q      Max
-0.24348   -0.03186    0.03740   0.07727   0.12669

Coefficients:
              Estimate Std. Error  t value Pr(>|t|)
(Intercept)  -8.4761     0.5007   -16.93  3.08e-10 ***
lnLength      3.4311     0.1330    25.80  1.49e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1229 on 13 degrees of freedom
Multiple R-squared:  0.9808,    Adjusted R-squared:  0.9794
F-statistic: 665.8 on 1 and 13 DF,  p-value: 1.495e-12
```

La retta che determina la regressione di y su x è determinata da due coefficienti. Il primo coefficiente, l'intercetta y , è il valore di y quando x è uguale a 0.

Il secondo coefficiente è la pendenza della retta o il cambiamento nel valore di y all'aumentare (o diminuire) di ogni unità di x .

Il software determina la retta di regressione attraverso il procedimento che minimizza la somma degli scarti dalla retta di regressione. Se la pendenza della retta è positiva, y cresce al crescere di x , se la retta è negativa y decresce all'aumentare di x .

La forza della relazione tra la variabile dipendente e la variabile indipendente è data dal coefficiente di correlazione, solitamente denominato r (o ρ). r varia da -1 (le variabili sono correlate negativamente) a 1 (perfetta relazione lineare) e un valore di r uguale a 0 indica la mancanza di relazione tra le due variabili (*paragrafo 4.7.1*).

Il coefficiente di determinazione di un modello di regressione lineare R^2 , che si ottiene semplicemente elevando r al quadrato, esprime la variabilità nella variabile dipendente spiegata dalla variabile indipendente. In parole più semplici, R^2 rappresenta la variazione nei valori di y che può essere giustificata dalla variazione di x .

Assumiamo che il termine di errore ε nel modello di regressione lineare sia indipendente da x e sia normalmente distribuito, con media 0 e varianza costante. Possiamo decidere se la relazione esistente tra x e y è statisticamente significativa testando, con il *test t di Student*, l'ipotesi nulla H_0 che $\beta_1=0$ (e che non esista quindi una relazione tra x e y).

Se il p-value < 0.05 possiamo rigettare l'ipotesi nulla e ciò significa che la relazione tra le variabili nel modello di regressione lineare è significativa.

9.2 Regressione lineare multipla

La *regressione lineare multipla* è una generalizzazione della retta di regressione semplice e ad essa sono applicati gli stessi vincoli sulla natura dei dati. La regressione lineare descrive la relazione tra due o più variabili indipendenti e una singola variabile dipendente.

Se abbiamo n osservazioni di una variabile dipendente y associate ai valori dei predittori x_1, \dots, x_n il modello di regressione lineare multipla può essere definito come segue:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon_i \text{ con } i = 1, \dots, n$$

dove β_i sono i parametri da stimare ed ε il termine d'errore.

Nel modello lineare semplice, l'inclinazione β_1 rappresenta la variazione che la variabile y presenta in corrispondenza di una variazione unitaria di x . Non si prende in considerazione nessun'altra variabile oltre all'unica variabile indipendente inclusa nel modello. Nel modello di *regressione multipla* l'inclinazione β_1 ci dice come varia y in corrispondenza di una variazione unitaria della variabile x_1 , quando si tiene conto anche degli effetti delle altre variabili.

Nel paragrafo precedente abbiamo visto che il coefficiente di determinazione consente di valutare la bontà del modello di regressione stimato. Nel modello di regressione multipla, dal momento che si è in presenza di almeno due variabili esplicative, il coefficiente di determinazione rappresenta la proporzione di variabilità della y spiegata dalle variabili esplicative.

Tuttavia, alcuni ricercatori ritengono che quando si ricorre a un modello di regressione multipla sia più opportuno utilizzare l'indice R^2 corretto (*adjusted R²*), che tiene conto anche del numero di variabili esplicative incluse nel modello e dell'ampiezza del campione. Il ricorso a questo tipo di indice si rende necessario soprattutto qualora si vogliano confrontare modelli di regressione che intendono spiegare la medesima variabile dipendente, impiegando un numero diverso di variabili esplicative.

Per eseguire correttamente un'analisi statistica è necessario analizzare ogni singola variabile dipendente in maniera bivariata, e quelle che risultano significativamente associate all'*outcome* sotto studio devono essere inserite in un modello multivariato per testare tale associazione, tenendo conto anche degli effetti delle altre variabili.

In generale nella costruzione di un modello il criterio principale da seguire è la parsimonia. Tale criterio impone di inserire in un modello il numero minimo di variabili indipendenti che consentano di spiegare la variabile risposta. I modelli di regressione con poche variabili esplicative sono di più facile interpretazione e sono meno esposti al rischio di *multicollinearità* (o correlazione) tra le variabili esplicative.

Un metodo di selezione del modello ampiamente usato è la *regressione stepwise*, che mira all'individuazione del modello "migliore" senza dover tuttavia considerare tutti i modelli possibili.

Nella *regressione stepwise*, infatti, a ciascuno stadio del processo di costruzione del modello le variabili esplicative possono essere sia aggiunte sia eliminate dal modello stesso. Il processo di selezione termina quando nessuna altra variabile può essere eliminata e nessuna può essere aggiunta.

9.3 Regressione logistica

La *regressione logistica* è una tecnica statistica usata quando si vuole stimare la probabilità di un evento dicotomico come ad esempio la presenza o l'assenza di una malattia o la morte o l'efficacia o meno di un trattamento.

La probabilità dell'evento è la variabile dipendente (y) e i vari fattori che la influenzano sono le variabili indipendenti (x_1, x_2, \dots, x_n), chiamati anche fattori di rischio.

Si potrebbe pensare di calcolare la probabilità che si verifichi l'evento in studio come una proporzione o una percentuale. Ad esempio: supponiamo che in 500 pazienti trattati con un particolare farmaco, 25 abbiano sviluppato un evento avverso, una controindicazione.

La proporzione di eventi avversi è $25/500$ o 0.05 o il 5% .

I risultati della regressione logistica sono presentati, più che in termini di proporzione di un evento, come "*odds*" dell'evento.

Nei *capitoli 2 e 4* abbiamo già parlato dell'*odds*: ricordiamone le principali caratteristiche essendo essenziale nel contesto della regressione logistica.

L'*odds*, termine difficilmente riproducibile efficacemente in italiano, ha una relazione diretta con la probabilità: l'*odds* è infatti il rapporto tra la probabilità che un evento si verifichi e la probabilità che l'evento non si verifichi.

$$odds = \frac{p}{1 - p}$$

dove p è la probabilità che l'evento accada.

Supponiamo di creare una tabellina 2x2 (**Tabella 9.2**) in cui vengono inseriti i dati riguardanti i pazienti esposti ad un determinato fattore di rischio e che hanno contratto una malattia in studio (a), pazienti esposti e che non hanno contratto la malattia in studio (b) e pazienti che non sono stati esposti al fattore di rischio e che hanno o non hanno sviluppato la malattia in studio (c e d , rispettivamente).

Tabella 9.2 Tabella che mostra i dati relativi ai pazienti esposti ad un fattore di rischio, evidenziandone lo stato di malattia o di assenza di malattia

		Malattia	
		Sì	No
Esposizione	Sì	a	b
	No	c	d

L'“*odds ratio*” o *OR* si calcola attraverso i semplici rapporti (*odds*) fra le frequenze osservate e non attraverso le proporzioni.

$$odds\ ratio = \frac{a/c}{b/d} = \frac{\frac{a}{b} \cdot d}{c} = \frac{a \cdot d}{b \cdot c}$$

Un valore di *OR* uguale a 1 indica che non esiste associazione tra malattia ed esposizione (l'*odds* di esposizione dei casi è identico all'*odds* di esposizione nei controlli).

Un valore di *OR* superiore ad 1 indica l'esistenza di un'associazione positiva tra esposizione e malattia mentre un valore inferiore a 1 indica un'associazione negativa (l'esposizione al fattore in studio protegge dalla malattia).

Il passo successivo è trasformare in “*logit*” gli *odds* della malattia. Il *logit* si calcola come logaritmo naturale (\ln) degli *odds*:

$$logit = \ln\left(\frac{p}{1 - p}\right) = \ln(odds)$$

Prima di dichiarare l'esistenza di un rapporto causa-effetto tra l'esposizione e la malattia, bisogna eseguire un test di significatività statistica (per escludere che la differenza sia dovuta al caso). Anche il calcolo degli intervalli di confidenza dell'*OR* (paragrafo 4.5.1) può darci l'idea della significatività dell'associazione tra rischio e malattia. Se l'intervallo di confidenza al 95% include l'unità, il rapporto causa-effetto tra l'esposizione e la malattia non è significativo.

Il massimo numero di *covariate* (paragrafo 9.4.1) che può essere inserito in un modello di *regressione logistica multipla* è strettamente dipendente dal numero di eventi e/o di casi di malattia. Una semplice regola è includere nel modello di regressione logistica multipla 1 *covariata* ogni 10 eventi. Per esempio: se abbiamo un campione di 1,000 individui e 20 casi di

malattia, il massimo numero di *covariate* da includere nel modello logistico multiplo dovrebbe essere pari a 2.

Come nella *regressione lineare*, anche nella *regressione logistica* il rapporto tra variabile dipendente ed indipendente è descritto da un'equazione:

$$\text{logit } y = \beta_0 + \beta_1 x_1$$

Per poter interpretare questa equazione, entrambi i membri devono essere elevati ad esponente e (il numero di Nepero).

β_0 è l'intercetta, ossia il valore del logaritmo naturale degli *odds* della malattia quando l'esposizione è zero.

β_1 è il coefficiente di regressione: indica di quanto aumenta in media il logaritmo naturale dell'*odds* della malattia nei pazienti esposti rispetto a quelli non esposti.

Il metodo per calcolare i coefficienti di regressione tiene conto di tutte le possibili combinazioni delle variabili indipendenti. I coefficienti di regressione sono calcolati con il *metodo della massima verosimiglianza*: viene massimizzata la probabilità che, per ogni dato individuo con una particolare combinazione di variabili indipendenti, l'*odds* dell'evento di interesse sarà simile all'*outcome* di tutti gli altri individui che possiedono la stessa combinazione di variabili indipendenti.

La formula generale dell'equazione di regressione logistica è simile a quella della regressione lineare multivariata, dove però come variabile dipendente viene usato il logaritmo dell'*odds* dell'*outcome*.

Per determinare la qualità del modello di *regressione logistica* devono essere valutate almeno tre proprietà fondamentali.

La prima proprietà è la calibrazione, definita come l'abilità di un modello di assegnare un appropriato rischio tra i pazienti su cui si simula il modello. La calibrazione dovrebbe essere dimostrata: qualitativamente da una tabella o un grafico che quantifichino l'*outcome* stimato rispetto all'*outcome* osservato nei gruppi di pazienti; quantitativamente tramite il *test di Hosmer-Lemeshow*.

La seconda proprietà per valutare l'accuratezza di un modello è la proprietà di "discriminazione", definita come l'abilità di un modello di distinguere tra quei pazienti che hanno o non hanno l'*outcome* di interesse.

Questa capacità del modello viene calcolata tramite la curva ROC (*receiver-operating characteristic*). Questa analisi è spesso utilizzata nei test di valutazione diagnostica ed è basata sui concetti di sensibilità e specificità e dal calcolo dell'area sotto la curva (c).

Una perfetta discriminazione è ottenuta con un valore di c pari a 1.

L'ultima e più importante proprietà per valutare l'accuratezza del modello è la validazione. La validazione viene definita come la dimostrazione che il valore predittivo del modello ottenuto nella popolazione in studio è simile se applicato ad un differente gruppo di pazienti. Il metodo più comune di validazione consiste nel costruire il modello su una parte della popolazione estratta casualmente (solitamente dal 20% al 50%: dipende dalla numerosità della popolazione in studio) e poi testare il modello ottenuto applicandolo sui pazienti non utilizzati per la costruzione del modello.

Box 9.1 - Regressione con R

Funzione	Descrizione
<code>modello <- lm(VarDip ~ Fattore, dati)</code>	Regressione lineare semplice.
<code>summary(modello)</code>	

Box 9.1 – Segue

<pre>modello <- lm(VarDip ~ Fattore1 + Fattore2, dati) summary(modello)</pre>	Regressione multipla. Abbiamo preso due fattori o variabili indipendenti, ma è solo un esempio: tutto dipende dal modello (lo stesso vale per le funzioni successive).
<pre>modello <- lm(VarDip ~ Fattore1*Fattore2, dati) summary(modello)</pre>	Regressione multipla con interazione fra fattori.
<pre>modello <- lm(cbind(VarDip1,VarDip2) ~ Fattore1 + Fattore 2, dati) summary(modello)</pre>	Regressione multivariata
<pre>modello <- glm(VarDip ~ Fattore1 + Fattore2, dati, family=binomial()) summary(modello)</pre>	Regressione logistica

9.4 Le varianti dell'ANOVA

Nel *paragrafo 4.6.3* abbiamo considerato l'*analisi della varianza* valutando gli effetti di una o più variabili indipendenti su una sola variabile dipendente. Ma la realtà in cui viviamo è molto più complessa: spesso vogliamo valutare più variabili dipendenti (es.: lo stato di benessere, il dolore percepito, i giorni di malattia per disabilità, la mortalità, etc.), che possono essere più o meno correlate fra loro. In altri casi, possiamo voler approfondire la varianza dovuta agli errori residui o a possibili fattori confondenti che la variabile indipendente non ci mostra (es.: gli effetti di un trattamento possono essere mediati da quanto una persona ha pagato quel trattamento).

Per svolgere queste analisi, l'ANOVA può essere espansa in *ANCOVA* (*analisi della covarianza*) per valutare i fattori confondenti, *MANOVA* (*analisi multivariata della varianza*) nel caso di più variabili dipendenti e *MANCOVA* (*analisi multivariata della covarianza*) nel caso di più variabili dipendenti e per valutare i fattori confondenti.

Vediamole adesso tutte e tre, senza entrare nel dettaglio dei calcoli matematici, e appoggiandoci a quanto detto sulla regressione per una spiegazione più agile.

9.4.1 Analisi della covarianza (ANCOVA)

L'*analisi della covarianza* o *ANCOVA* serve per ridurre nella classica ANOVA gli effetti di eventuali variabili, dette *covariate*, non esplicitamente incluse nel disegno di studio. Come possiamo ricordarci (*paragrafo 4.6.3*), nell'ANOVA vi è sempre una componente di errore o residuo o *variabilità non spiegata* negli effetti che la variabile indipendente ha sulla variabile dipendente.

Parte di questo errore non spiegato è dovuto a variabili confondenti esterne al quesito principale dello studio, oppure a differenze within-subjects. Se noi consideriamo questi aspetti possiamo ridurre la componente di errore non spiegata e quindi avere un'analisi migliore degli effetti della variabile indipendente su quella dipendente.

Un classico esempio per cui viene usata l'*ANCOVA* è l'ANOVA a misure ripetute quando il fattore within-subjects è il tempo (*paragrafi 4.6.3.3-4*). Il ragionamento è il seguente: se studiamo gli effetti di un trattamento, il valore finale della variabile risposta

non dipende solo dal trattamento, bensì anche dal valore iniziale della variabile risposta stessa prima del trattamento (es.: il dolore finale percepito dipende anche da qual era il dolore di partenza). In tal caso, il valore della variabile dipendente pre-trattamento è la *covariata* che media/influenza gli effetti del trattamento, ossia del fattore studiato. Lo stesso ragionamento si può applicare in caso di studi più complessi (es.: con T0, T1, T2, etc.).

Altre *covariate* sono tutti quegli elementi che possono influenzare l'effetto del trattamento: l'età delle persone, il sesso, etc. Infatti ogni persona può rispondere diversamente ad uno "stesso" trattamento a seconda delle sue caratteristiche, che è bene considerare per poter valutare al meglio l'efficacia del trattamento, così da sapere al meglio quando e a chi "applicarlo".

LANCOVA deve rispettare le stesse assunzioni dell'ANOVA, ossia l'indipendenza delle osservazioni svolte, la Normalità della distribuzione dei dati e l'omogeneità delle varianze, oltre ad altre assunzioni sue specifiche:

- la *covariata* deve essere una variabile numerica continua (in realtà può essere anche categorica, ma in questo caso diversi autori preferiscono non parlare di ANCOVA);
- per ogni variabile indipendente, la relazione fra la *covariata* e la variabile dipendente deve essere lineare (valutabile tramite il coefficiente di regressione lineare);
- nei diversi gruppi (o per ogni livello della variabile indipendente), le curve di regressione fra la *covariata* e la variabile dipendente devono essere omogenee, ossia tendenzialmente parallele fra loro (il parametro β_i di ogni linea di regressione deve essere il medesimo o, quantomeno, simile). In caso contrario, la *covariata* influenza la variabile dipendente diversamente nei vari gruppi: in un gruppo può ad esempio aumentare il valore della variabile risposta (correlazione positiva fra *covariata* e variabile risposta), mentre in un altro diminuirli (correlazione negativa).

Questa situazione non necessariamente è un evento negativo, anzi in alcuni casi può essere anche interessante (es.: in uno studio multicentrico su diversi ospedali, l'esperienza dei terapeuti può essere molto diversa, così come i pazienti visti, e queste sono *covariate* che mediano gli effetti del trattamento), ma deve essere valutato in maniera specifica tramite modelli lineari multilivello;

- la *covariata* deve essere indipendente da ogni variabile indipendente, altrimenti gli effetti della variabile indipendente potranno essere spuri e non affidabili. Se i gruppi sotto studio differiscono per la *covariata*, fatto che succede quando i partecipanti non sono ad esempio randomizzati così da creare gruppi omogenei, considerare la *covariata* non bilancerà le differenze fra i gruppi, ma anzi potrà diminuirle o aumentarle a seconda del caso. Pertanto, la *covariata* non deve essere un fattore che distingue i gruppi sotto studio fra loro (per approfondimenti, Miller & Chapman, 2001). In tal caso, prima di usare l'ANCOVA, potremmo eseguire dei test t o un'ANOVA per valutare se i gruppi differiscono per la *covariata*: in caso di esito non significativo, possiamo usare l'ANCOVA con quella *covariata* con maggior tranquillità (Field *et al.*, 2012).

Questa assunzione vale principalmente per gli studi sperimentali, dove i ricercatori manipolano le variabili. Infatti, più propriamente le *covariate* devono essere indipendenti dai trattamenti sperimentali, altrimenti risulta difficile comprendere bene gli effetti del trattamento. Negli studi osservazionali, invece, dove la randomizzazione non avviene, spesso *covariata* e variabile indipendente non sono indipendenti, ma ci troviamo in un altro disegno di studio, che va interpretato diversamente e in cui la rimozione della *covariata* può alterare le stime dell'effetto della variabile indipendente su quella dipendente (Grace-Martin, 2012)

Francesco Cerritelli • Diego Lanaro

Elementi di ricerca in Osteopatia e Terapie Manuali



www.edises.it



€ 35,00