

Alan Agresti
Barbara Finlay

Statistica per le scienze sociali

Edizione italiana a cura di
Mariano Porcu
Università di Cagliari



© 2009 Pearson Paravia Bruno Mondadori S.p.A.

Authorized translation from the English language edition, entitled: STATISTICAL METHODS FOR THE SOCIAL SCIENCE, 4TH edition, by ALAN AGRESTI; BARBARA FINLAY, published by Pearson Education, Inc, publishing as Prentice Hall, Copyright © 2009.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

Italian language edition published by Pearson Paravia Bruno Mondadori S.p.A., Copyright © 2009.

Le informazioni contenute in questo libro sono state verificate e documentate con la massima cura possibile. Nessuna responsabilità derivante dal loro utilizzo potrà venire imputata agli Autori, a Pearson Paravia Bruno Mondadori S.p.A. o a ogni persona e società coinvolta nella creazione, produzione e distribuzione di questo libro.

Per i passi antologici, per le citazioni, per le riproduzioni grafiche, cartografiche e fotografiche appartenenti alla proprietà di terzi, inseriti in quest'opera, l'editore è a disposizione degli aventi diritto non potuti reperire nonché per eventuali non volute omissioni e/o errori di attribuzione nei riferimenti.

I diritti di riproduzione e di memorizzazione elettronica totale e parziale con qualsiasi mezzo, compresi i microfilm e le copie fotostatiche, sono riservati per tutti i paesi.

LA FOTOCOPIATURA DEI LIBRI È UN REATO. Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633.

Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da AIDRO, corso di Porta Romana n. 108, 20122 Milano, e-mail segreteria@aidro.org e sito web www.aidro.org.

Curatore per l'edizione italiana: Mariano Porcu

Traduzione: Vito Muggeo, Mariano Porcu e Nicola Tedesco

Copy-editing: Roberto Molteni

Impaginazione: M.T.M. s.n.c. Fotocomposizione

Grafica di copertina: Nicolò Cannizzaro

Stampa: Arti Grafiche Battaia S.&C.- Zibido S. Giacomo (MI)

Tutti i marchi citati nel testo sono di proprietà dei loro detentori.

978-88-7192-538-7

Printed in Italy

1^a edizione: gennaio 2009

Ristampa	Anno
00 01 02 03 04	09 10 11 12 13

Ai miei genitori Louis J. Agresti e Marjorie H. Agresti

Indice generale

Prefazione all'edizione italiana		XI
Prefazione		XIII
Capitolo 1	Introduzione	1
	1.1 Introduzione alla metodologia statistica	1
	1.2 Statistica descrittiva e statistica inferenziale	4
	1.3 Il ruolo dei computer nella statistica	6
	1.4 Riassunto del capitolo	8
	PROBLEMI	8
Capitolo 2	Campionamento e misurazione	11
	2.1 Le variabili e la loro misurazione	11
	2.2 Casualizzazione	15
	2.3 Variabilità campionaria e potenziali fattori di distorsione	19
	2.4 Altri metodi di campionamento probabilistici	23
	2.5 Riassunto del capitolo	27
	PROBLEMI	28
Capitolo 3	Statistiche descrittive	33
	3.1 Descrivere i dati con tabelle e grafici	33
	3.2 Descrivere il "centro" dei dati	41
	3.3 Descrivere la variabilità dei dati	48
	3.4 Misure di posizione	53
	3.5 Statistiche descrittive bivariate	58
	3.6 Statistiche campionarie e parametri della popolazione	61
	3.7 Riassunto del capitolo	62
	PROBLEMI	64
Capitolo 4	Distribuzioni di probabilità	75
	4.1 Introduzione alla probabilità	75
	4.2 Distribuzioni di probabilità per variabili discrete e continue	77
	4.3 La distribuzione di probabilità normale	81
	4.4 Le distribuzioni campionarie descrivono come variano le statistiche	88

	4.5	Distribuzioni campionarie delle medie campionarie	92
	4.6	Richiami su: popolazione, dati campionari e distribuzioni campionarie	98
	4.7	Riassunto del capitolo	102
		PROBLEMI	103
Capitolo 5		Inferenza statistica: la stima	111
	5.1	Stima puntuale e intervallare	111
	5.2	Intervallo di confidenza per una proporzione	115
	5.3	Intervallo di confidenza per la media	121
	5.4	Scelta della dimensione campionaria	128
	5.5	Intervalli di confidenza per la mediana e altri parametri	135
	5.6	Riassunto del capitolo	137
		PROBLEMI	139
Capitolo 6		Inferenza statistica: test di significatività	147
	6.1	Le cinque parti di un test di significatività	148
	6.2	Test di significatività per una media	151
	6.3	Test di significatività per una proporzione	160
	6.4	Decisioni e tipi di errore nelle verifiche di ipotesi	163
	6.5	Limiti dei test di significatività	167
	6.6	Calcolare la P (errore del II tipo)	170
	6.7	Test per una proporzione nel caso di piccoli campioni - La distribuzione binomiale	173
	6.8	Riassunto del capitolo	177
		PROBLEMI	179
Capitolo 7		Confronto tra due gruppi	187
	7.1	Introduzione al confronto tra gruppi	187
	7.2	Dati categoriali: confronto fra due proporzioni	191
	7.3	Dati quantitativi: confronto di due medie	195
	7.4	Confronto di medie per campioni dipendenti	197
	7.5	Altri metodi per il confronto di medie	201
	7.6	Altri metodi per il confronto fra proporzioni	206
	7.7	Statistiche non parametriche per il confronto fra gruppi	209
	7.8	Riassunto del capitolo	213
		PROBLEMI	214
Capitolo 8		L'analisi dell'associazione tra variabili categoriali	225
	8.1	Tabelle di contingenza	225
	8.2	Test chi-quadro di indipendenza	228
	8.3	I residui: individuare la struttura dell'associazione	233
	8.4	Misure di associazione in tabelle di contingenza	237
	8.5	Associazione tra variabili ordinali	243
	8.6	Inferenza per le misure ordinali di associazione	247
	8.7	Riassunto del capitolo	250
		PROBLEMI	251

Capitolo 9	Regressione lineare e correlazione	259
	9.1 Relazioni lineari	260
	9.2 Equazione di previsione dei minimi quadrati	263
	9.3 Il modello di regressione lineare	269
	9.4 Misura dell'associazione lineare: la correlazione	273
	9.5 Inferenza per il coefficiente angolare e la correlazione	279
	9.6 Assunzioni del modello e violazioni delle stesse	287
	9.7 Riassunto del capitolo	291
	PROBLEMI	293
Capitolo 10	Introduzione alle relazioni multivariate	305
	10.1 Associazione e causalità	305
	10.2 Controllo per altre variabili	308
	10.3 Tipi di relazioni multivariate	311
	10.4 Aspetti inferenziali	317
	10.5 Riassunto del capitolo	318
	PROBLEMI	319
	Appendice A - SPSS e SAS per l'analisi statistica	325
	Appendice B - Risposte per una selezione di problemi con numerazione dispari	335
	Tavole	351
	Bibliografia	355
	Indice analitico	357

Prefazione all'edizione italiana

Statistical Methods for the Social Sciences è un libro con una storia di quasi tre decenni. Ha avuto un'ampissima diffusione come strumento di divulgazione delle conoscenze statistiche presso il vasto pubblico degli studenti di scienze sociali: è stato adottato in oltre 170 istituzioni universitarie statunitensi (fra le quali alcune della famosa Ivy League) e, nella sua terza edizione in lingua inglese, ha avuto una diffusione superiore alle 50 000 copie. Alla base di questa fortunata storia editoriale vi è la sostanza di un libro che riesce a presentare i metodi statistici di base enfatizzando il loro uso per le concrete finalità conoscitive: la spiegazione dei metodi è sempre corredata con il supporto di numerosi esempi tratti dall'attualità e dai contesti applicativi propri delle scienze sociali, esempi che rendono l'apprendimento interessante e offrono al lettore un affascinante punto di osservazione statistico su eventi recenti che hanno interessato o interessano l'opinione pubblica mondiale.

Pur con questa spiccata connotazione "pratica" il testo non rinuncia a una trattazione precisa degli aspetti metodologici più rilevanti non entrando, tuttavia, nel dettaglio della loro formalizzazione matematica. Tra gli autori, Alan Agresti, curatore da solo delle due più recenti edizioni, è un eminente studioso che concentra i suoi interessi di ricerca nell'ambito dell'analisi dei dati discreti, della biostatistica, della statistica sociale oltre che un esperto della cosiddetta *statistical education*, cioè della presentazione dei concetti statistici ai non-statistici. È, inoltre, un ricercatore molto noto nell'ambito della trattazione dei dati categoriali, quel tipo di dati, cioè, che sono il principale materiale statistico che si incontra nella ricerca sociale.

La traduzione di un testo è un'operazione delicata. In accordo con l'editore si è deciso di adottare una traduzione, ancorché non letterale, strettamente fedele, però, allo spirito dell'opera in madrelingua senza alcuna contestualizzazione all'ambito italiano¹. Le applicazioni e i problemi proposti, gioco forza, risulteranno, quindi, molto "statunitensi". Crediamo che questo abbia permesso di conservare quasi intatto lo spirito che impronta il libro di Agresti (oltreché, molti dei suoi lavori metodologici), un autore che ha saputo proporre esempi divenuti spesso dei "classici" didattici².

L'edizione originale in lingua inglese comprende materiale didattico che, nelle intenzioni dell'autore, necessiterebbe, per essere adeguatamente sviluppato, di due corsi di insegnamento di durata semestrale. Vi sono, infatti, 6 capitoli dedicati alla trattazione di metodologie statistiche avanzate tra le quali l'analisi della varianza, l'analisi di regressione multipla e la regressione logistica. Considerando il pubblico di riferimento principale dell'edizione italiana, cioè gli studenti delle lauree triennali nei quali viene previsto un unico insegnamento semestrale di statistica, si è deciso di tradurre solo i primi 10 capitoli.

Approfitto di questa introduzione per ricordare allo studente che scrivere, tradurre e pubblicare un libro è un'operazione complessa e costosa che si basa sul lavoro di molte persone. Fotocopiare un libro è un'operazione illegale che oltre a danneggiare materialmente tutte le persone che lavorano nell'editoria, minaccia la sopravvivenza stessa del libro come lo strumento più importante per la trasmissione del sapere scientifico.

¹ Per motivi di praticità, in ragione dei numerosi riferimenti del testo a risultati di analisi svolte con l'ausilio di software, si è deciso di mantenere il punto come segno di separazione per le cifre decimali.

² Ad esempio, quelli sulla pena di morte per bianchi e neri.

Desidero infine ringraziare Vito Muggeo e Nicola Tedesco per il loro supporto e per la loro indispensabile collaborazione³. Un immenso grazie per la loro infinita pazienza va, in particolare, ad Anna e a Francesco.

*Mariano Porcu
Università di Cagliari*

Supplementi on-line

All'indirizzo www.prenhall.com/agresti è disponibile per i docenti che adottano il testo, protetto da password, l'Instructor's Resource Manual con le soluzioni degli esercizi presenti nel volume (cliccare sulla copertina e accedere all'area Resources).

Allo stesso indirizzo, e liberamente accessibili, sono disponibili per gli studenti le applet e i data set. Il materiale è in lingua inglese.

Le slide Power Point in italiano, per i docenti che adottano il testo, sono disponibili invece all'indirizzo <http://hpe.pearson.it/agresti>.

Ringraziamenti

L'editore ringrazia la professoressa Alessandra Petrucci, dell'Università di Firenze, per il prezioso contributo all'opera con i suoi consigli e suggerimenti.

³ V. Muggeo ha tradotto i Capitoli 7, 8, 9 e parte del 10; N. Tedesco i Capitoli 4, 5, 6 e parte del 10. I Capitoli 1, 2 e 3 sono stati tradotti da M. Porcu.

Prefazione

Quasi tre decenni fa, quando è stata iniziata la stesura della prima edizione di questo libro, l'obiettivo che ci si era prefissi era quello di presentare i metodi statistici con uno stile che enfatizzasse i concetti e le loro applicazioni in ambito sociale. Per cercare di pervenire all'obiettivo si era deciso di privilegiare le spiegazioni su come i metodi dovevano essere usati e interpretati piuttosto che sulle loro dimostrazioni teoriche e sugli aspetti matematici e computazionali.

Questa quarta edizione del libro pone un'enfasi ancora maggiore sui concetti e sulle loro applicazioni con una notevole attenzione, sia negli esempi sia negli esercizi, ai cosiddetti *dati reali*. Si è, quindi, deciso di continuare a trattare gli aspetti matematici con poco rilievo, specie nel trattare il tema della probabilità che risulta essere, talvolta, l'ostacolo su cui inciampano gli studenti nel loro percorso di apprendimento. Allo stesso tempo, questo testo non vuole essere un "ricettario" per cucinare le informazioni quantitative: semplificare eccessivamente la trattazione dei metodi sposando, appunto, un approccio stile ricettario, non porta a una buona pratica della statistica.

Cambiamenti della quarta edizione

Da quando è stata pubblicata la prima edizione di questo testo, l'aumento della capacità di calcolo informatico e la simultanea maggiore facilità con cui si accede all'impiego di software statistici ha avuto un rilevante impatto sulle modalità con le quali gli studiosi di scienze sociali analizzano i dati quantitativi. In ragione di ciò, in questo libro non verranno presentate le formule che semplificano lo svolgimento manuale dei calcoli e, inoltre, metodi computazionalmente complessi, quali ad esempio la regressione, verranno presentati enfatizzando l'interpretazione degli output ottenuti attraverso i software piuttosto che esponendo nel dettaglio come debbano essere svolti i loro calcoli. Il testo contiene numerosi esempi di output di analisi svolte attraverso l'impiego di software (principalmente SPSS e SAS) sia all'interno dei capitoli sia nelle parti dedicate ai problemi. In questa edizione è compresa, inoltre, un'appendice nella quale viene illustrato come utilizzare i software SPSS e SAS per svolgere le analisi presentate in ciascun capitolo; il sito Web contiene poi alcuni link per ottenere informazioni su altri software statistici.

La presentazione di esempi semplici ma basati su dati reali oltre che la proposizione di numerosi problemi per lo studio individuale è fondamentale per l'apprendimento dello studente. In questa edizione, dando maggiore enfasi ai dati reali, sono stati aggiunti numerosi nuovi problemi e sono stati aggiornati molti di quelli proposti nelle precedenti edizioni. La parte Problemi di ciascun capitolo è stata divisa in due gruppi. Il primo, chiamato *Concetti di base*, contiene problemi direttamente collegati ai metodi trattati nel testo. Il secondo, chiamato *Applicazioni*, contiene problemi riferiti ad analisi di dati da completare oppure finalizzati alla comprensione di concetti fondamentali. I grandi dataset a cui si fa riferimento negli esempi e nei problemi e fra i quali è compreso quello presentato nel Problema 1.11 e che viene riproposto nella parte Problemi di ciascun capitolo è disponibile all'indirizzo Internet

<http://www.stat.ufl.edu/~aa/social/data.html>

Questa edizione, rispetto alle precedenti, contiene diversi cambiamenti e integrazioni che vanno nella direzione di un approccio più moderno alle tematiche trattate.

- Viene data maggiore enfasi agli esempi basati su **dati reali** e all'interpretazione dei risultati ottenuti attraverso il **software statistico**. Allo scopo di aiutare lo studente ad apprendere i concetti fondamentali sulle distribuzioni campionarie, sugli intervalli di confidenza e sui test di significatività sono presenti problemi in cui si chiede agli studenti di utilizzare alcuni applicativi disponibili nel sito web predisposto dall'editore.
- Il Capitolo 3, dedicato alla statistica descrittiva, contiene un paragrafo specificamente dedicato, fra le misure di posizione, ai percentili e ad altri argomenti a essi collegati quali i box plot e le osservazioni anomale. Allo scopo di dare una prima idea allo studente su tabelle di contingenza e analisi di regressione, cioè, sull'importanza di prendere in considerazione più di una variabile per volta, sempre nel Capitolo 3 è contenuta una breve introduzione ai metodi descrittivi per le analisi bivariate. In tale introduzione sono presentati i concetti di associazione statistica oltre a quelli di variabile risposta e di variabile esplicativa.
- Il Capitolo 4 contiene un nuovo paragrafo in cui viene presentato il concetto di frequenza relativa e in cui vengono riassunte, in breve, tre regole base sulla probabilità che verranno applicate in seguito nel testo.
- L'argomento dell'inferenza per le medie dei Capitoli 5, 6 e 7 è stato completamente riformulato in modo tale che lo stesso si basi interamente sulla distribuzione t piuttosto che sulla distribuzione z sia per i piccoli sia per i grandi campioni. Ciò allo scopo di rendere i risultati confrontabili con quelli ottenuti attraverso i software statistici. Nella presentazione dei metodi si è enfatizzato che l'assunzione di normalità per la distribuzione t è necessaria principalmente per l'inferenza unilaterale con piccoli campioni.
- In parte in ragione dei cambiamenti determinati dall'aver scelto di usare sempre la distribuzione t per fare inferenza sulla media, nel Capitolo 5 gli intervalli di confidenza per una proporzione (che si basano sulla distribuzione normale) sono stati presentati prima di quelli per la media. In tal modo lo studente potrà apprendere i concetti di base sugli intervalli di confidenza utilizzando immediatamente le nozioni sulla distribuzione campionaria normale che ha appreso alla fine del Capitolo 4. Questo spostamento in avanti del paragrafo dedicato alla distribuzione t dovrebbe aver ridotto il rischio che lo studente possa confondersi nell'affrontare troppi nuovi argomenti tutti in una volta.
- Il Capitolo 7 sul confronto fra due gruppi contiene un nuovo paragrafo dedicato ai concetti dell'analisi bivariata in cui viene ricordata agli studenti la distinzione fra variabili risposta e variabili esplicative, vengono definiti i campioni indipendenti e dipendenti, viene discusso come confrontare due gruppi attraverso le differenze o i rapporti fra i valori di due parametri e viene proposta una formula generale per trovare l'errore standard della differenza fra due stime indipendenti. Viene poi introdotto il concetto di modello statistico.

A prescindere dal tipo di dati che viene analizzato, si ritiene che un approccio di analisi basato su un qualche paradigma modellistico in cui si enfatizza la stima dei parametri sia più utile dell'approccio, adottato da molti manuali di statistica, basato sui test di ipotesi. Infatti, i capitoli di base sull'inferenza (5-8) spiegano quali siano i vantaggi che ha l'uso degli intervalli di confidenza rispetto ai test di significatività. Non volendo essere un ricettario enciclopedico ma, piuttosto uno strumento didattico, questo libro non tenta di presentare una completa rassegna di tutti i metodi statistici disponibili.

Ringraziamenti

Desidero ringraziare Barbara Finlay per il suo contributo nella stesura delle prime due edizioni di questo libro. Spero che la combinazione tra i nostri rispettivi campi di esperienza sia stata in grado di produrre un testo che è uno strumento utile sia per l'apprendimento della statistica in generale sia per i metodi della ricerca sociale.

Ringrazio tutti coloro che hanno investito un rilevante ammontare del loro tempo nell'aiutarmi ad aumentare la fruibilità del libro. Grazie a Sarah Streett per aver aggiornato i dati di molti esempi e problemi, a Jackie Miller per il suo eccellente lavoro di revisione e a Traci Douglas per il suo lavoro di Production Editor. Grazie a Arne Bathke, David Most, Youqin Huang e Michael Lacy per i loro commenti a questa edizione. Altre persone che hanno collaborato per i data set sono Roslynn Brain, Beth Chance, Brent Coull, Alfred DeMaris, Mary Gray, Brian Gridley, Ralitzia Gueorguieva, John Henretta, Ira Horowitz, Harry Khamis, Bernhard Klingenberg, Jacalyn Levine, Michael Radelet, Paula Rausch, Euijung Ryu, Paul Smith, Robert Wilson e Zoe Ziliak. Grazie anche a tutti coloro che con i loro commenti hanno collaborato alla preparazione delle prime tre edizioni, in special modo a Burke Grandjean, Susan Reiland, Maureen Hallinan, Shirley Scritchfield, Sonja Wright, Douglas Zahn, Jeff Witmer, E. Jacquelin Dietz, Dorothy K. Davidson e Mary Sue Younger. Ringrazio i miei editori di questa e delle precedenti edizioni Petra Recter e Ann Heath della Prentice Hall per il loro costante supporto e incoraggiamento.

Infine, uno speciale ringraziamento va a mia moglie Jacki Levine per il suo supporto nel lavoro di editing della terza edizione oltre che per i suoi incoraggiamenti nella preparazione della presente edizione.

Alan Agresti

Gainesville, Florida