

Barbara Pacini

Lucio Picci

Introduzione alla Statistica



INDICE

	<i>pag.</i>
Prefazione	11
Capitolo 1 – <i>Che cosa è la statistica?</i>	15
1.1 Introduzione	15
1.2 Lo studio dei fenomeni collettivi	16
1.3 La descrizione dei dati, o “statistica descrittiva”	17
1.4 La probabilità	18
1.5 Il campionamento e l’“inferenza statistica”	19
1.6 Relazioni tra fenomeni e le previsioni	20
1.7 Decisioni in regime di incertezza	21
1.8 Le fonti di informazione statistica	21
Capitolo 2 – <i>La statistica descrittiva: riassumere le informazioni</i>	25
2.1 Introduzione	25
2.2 Tipi di dati	26
2.3 Scale di misura	27
2.4 Le rappresentazioni tabellari	29
2.5 Serie storiche e serie territoriali	34
2.6 I dati raggruppati in classi	35
2.7 Le rappresentazioni grafiche	37
2.8 Misure di tendenza centrale dei dati	41
2.9 Misure di variabilità dei dati	51
2.10 I rapporti statistici	61
Capitolo 3 – <i>La statistica descrittiva: il caso bivariato</i>	69
3.1 Introduzione	69
3.2 Le tabelle a doppia entrata	69
3.3 Le misure di associazione: l’indice χ^2	73
3.4 La correlazione tra caratteri quantitativi	77
3.5 Il questionario	80

Capitolo 4 – <i>La probabilità</i>	83
4.1 Introduzione	83
4.2 Gli eventi	84
4.3 Il Diagramma di Venn	84
4.4 La probabilità	85
4.5 La regola dell’addizione	88
4.6 La probabilità condizionata	89
4.7 L’indipendenza tra eventi	91
4.8 La regola della moltiplicazione	92
4.9 Il Teorema di Bayes	93
 Capitolo 5 – <i>Variabili casuali e distribuzioni di probabilità</i>	 99
5.1 Introduzione	99
5.2 Variabili casuali	99
5.3 Distribuzione di probabilità per variabili discrete	100
5.4 Valore atteso di una variabile casuale discreta, o aspettativa matematica	101
5.5 Varianza di una variabile casuale discreta	102
5.6 La distribuzione binomiale	103
5.7 La distribuzione ipergeometrica	108
5.8 Distribuzione di probabilità per variabili continue	110
5.9 La distribuzione normale	113
5.10 La standardizzazione della distribuzione normale	116
5.11 L’utilizzo delle tavole della distribuzione normale standardizzata ...	118
5.12 La distribuzione normale come approssimazione della distribuzione binomiale	120
5.13 La distribuzione chi-quadro χ^2 : uso delle tavole	121
5.14 La distribuzione t di Student: uso delle tavole	122
5.15 Altre distribuzioni	124
 Capitolo 6 – <i>L’analisi campionaria</i>	 127
6.1 Introduzione	127
6.2 Il campionamento casuale	128
6.3 La distribuzione campionaria	129
6.4 La distribuzione campionaria della media campionaria	135
6.5 Il teorema del limite centrale: una illustrazione	140
6.6 La distribuzione campionaria di una frequenza relativa	144
 Capitolo 7 – <i>La stima</i>	 147
7.1 Introduzione	147
7.2 Proprietà degli stimatori	149

7.3 Alcuni stimatori	151
7.4 Metodi di derivazione degli stimatori	152
7.5 Stima per intervallo della media della popolazione	153
7.6 Stima per intervallo della frequenza relativa della popolazione	158
Capitolo 8 – <i>Verifica di ipotesi</i>	161
8.1 Introduzione	161
8.2 Verifica di ipotesi: i concetti di base	162
8.3 Verifica di ipotesi per la media	166
8.4 Verifica di ipotesi per una frequenza relativa	171
8.5 La potenza di un test	174
8.6 Verifica di ipotesi: altri casi	182
Capitolo 9 – <i>Test sulla varianza e l'analisi della varianza</i>	187
9.1 Introduzione	187
9.2 La distribuzione della varianza campionaria	188
9.3 La distribuzione del χ^2 e la distribuzione della varianza campionaria .	191
9.4 Il test del χ^2	194
9.5 La verifica di ipotesi sull'uguaglianza di varianze e la distribuzione F	197
9.6 L'analisi della varianza	200
Capitolo 10 – <i>Il modello di regressione lineare</i>	209
10.1 Introduzione	209
10.2 La covarianza e la correlazione	209
10.3 Il modello di regressione lineare	215
10.4 La stima dei parametri della retta di regressione	217
10.5 Le caratteristiche dello stimatore dei minimi quadrati	221
10.6 L'errore standard della stima	221
10.7 L'indice di determinazione lineare	223
10.8 La verifica di ipotesi	225
Appendice A – <i>L'operatore sommatoria</i>	229
Appendice B – <i>Tavole statistiche</i>	231
Aree sottese alla distribuzione normale standardizzata	231
Valori della distribuzione t di Student	232
Valori della distribuzione χ^2	233
Valori della distribuzione F	234
Indice analitico	243

*“Compulsory statistics courses have received a bad press.
Indeed, on the typical college campus, enrollment in such a course
ranks somewhere in popularity between laws establishing
a minimum drinking age and mandatory draft registration.
In part, this reputation is deserved.”*

Paul Newbold

*“Lustrum post annum alterum et quadragensimum fec[i].
Quo lustrum civium Romanorum censa sunt capita
quadragiens centum millia et sexag[i]nta tria millia.”*

Ottaviano Augusto, Res Gestae, VIII

Prefazione

Esiste un neologismo anglosassone, “innumeracy”, che in italiano potremmo scherzosamente tradurre come “snumeratezza”, per indicare le difficoltà che molti incontrano nel comprendere e nel trattare in modo corretto le informazioni di tipo quantitativo. La lettura dei giornali, o una visione anche distratta di molti programmi televisivi, mostrano come questa incapacità sia assai diffusa: siamo quotidianamente vittime di percentuali che non sommano a cento e di grafici artisticamente costruiti che però non dicono nulla, e ci tocca di sopportare, a ogni periodica pubblicazione dei dati sul livello dei prezzi, affermazioni come “è diminuito il costo della vita”, per indicare invece che il costo della vita è sì aumentato, ma meno che in precedenza.

Si potrebbe persino ironizzare su quanto sia difficile il mestiere dello statista, in un mondo ingrato e crudele in cui le statistiche vengono quotidianamente violentate. Ma il punto evidentemente è un altro: viviamo in un ambiente in cui la quantità di informazione disponibile è sempre più grande, e sempre maggiore è la porzione di questa informazione che si presenta in forma quantitativa: dati e statistiche ufficiali, sondaggi d’opinione, tabelle, grafici, eccetera. Non riuscire ad assimilare e ad elaborare questi dati in modo adeguato significa rinunciare a utilizzare una parte sempre più rilevante dell’informazione disponibile. Il ragionamento vale ovviamente per la nostra vita professionale, presente o futura, ma non solo, perché comprendere in modo adeguato le informazioni in nostro possesso è cruciale anche per potere esercitare al meglio i nostri diritti e doveri di cittadini.

Questi sono i motivi principali che giustificano la presenza di un corso di introduzione alla statistica all’interno di numerosi corsi di laurea. La statistica, in senso lato, è quella disciplina che si occupa dell’analisi dei dati, principalmente dei dati di tipo quantitativo; una descrizione più compiuta degli argomenti di cui

si interessa la statistica è fornita nel primo capitolo del libro. Questo è un libro breve, che tenta di occuparsi del nocciolo dei problemi e che per questo motivo omette molti dettagli e sviluppi anche importanti. Nella sua natura essenzialmente non enciclopedica, questo testo sarà forse utile anche a chi, se pure per la prima volta e al di fuori di un corso di studi, desidera avvicinarsi allo studio delle tecniche di base della statistica.

L'accento sulle tecniche di base non deve però trarre in inganno: in esse si trovano molte delle idee portanti che costituiscono l'oggetto delle procedure statistiche più avanzate. Se questo libro coglie nel segno, chi lo utilizzerà apprenderà, almeno a un livello intuitivo, non soltanto l'utilizzo di certe tecniche, ma anche molti dei problemi di fondo dell'analisi statistica in generale.

Per ogni argomento, numerosi problemi risolti permetteranno al lettore di verificarne la comprensione; chi desidera esercitarsi ulteriormente potrà consultare uno dei tanti libri appositi, tra i quali segnaliamo, di Lucio Picci, "Esercizi di statistica", Clueb, Bologna, 2001, che costituisce un accompagnamento a questo testo.

Paul Newbold, nella citazione riportata in apertura, ci ricorda che i corsi universitari obbligatori di statistica hanno una cattiva reputazione presso gli studenti. Sarebbe pretenzioso, per noi, sperare di cambiare una tale situazione. Alla legittima antipatia verso i numeri, le percentuali e le probabilità, ci accontenteremo di riuscire ad affiancare un punto di vista diverso. Contare, enumerare, riassumere e utilizzare opportunamente l'informazione quantitativa è sempre stata un'attività così importante, da assumere spesso dei significati rituali profondi. Nella lingua latina il censimento, che avveniva ogni cinque anni, era seguito dal "lustrum", una cerimonia di purificazione in cui un maiale, una pecora e un toro venivano sacrificati. L'identificazione tra il conteggio e il rito era così forte, che il nome del secondo, come nella citazione del Res Gestae di Augusto, finì per identificare il primo, e ancora oggi il lustro è, per noi, un periodo di cinque anni.

L'auspicio è che il lettore di questo libro affronti la statistica se non con favore, almeno con la consapevolezza che essa studia delle attività che da sempre rivestono un ruolo importante all'interno della società; un ruolo così importante da avere quasi sempre assunto un significato ritualistico, visibile non soltanto nelle antiche cerimonie, ma anche oggi, nell'aurea quasi sacrale che i grafici, i sondaggi d'opinione e le indagini di ogni genere assumono nei mezzi di informazione e nel discorso pubblico.

Questo libro costituisce una versione rivista e sensibilmente ampliata di un testo pubblicato nel 1998, con Lucio Picci come unico autore. La presente edizione nasce dall'esperienza comune di insegnamento degli autori presso la Facoltà di Scienze Politiche e di Economia dell'Università di Bologna. Gli stu-

denti, nel corso degli anni, hanno svolto un ruolo essenziale per permettere di trasformare delle "dispense", nate quasi casualmente, in un testo compiuto. Guido Gambetta, Roberto Golinelli, Corrado Lagazio, Fabrizia Mealli e Monica Pratesi hanno fornito consigli assai utili e hanno permesso di migliorare il testo originale. Margherita Borri, Luca Fanelli e Rita Romeo hanno corretto alcuni errori e hanno fornito molti utili commenti. A tutti loro va il nostro ringraziamento.

Barbara Pacini e Lucio Picci

Bologna, settembre 2000