

CURRICULUM DELL'ATTIVITÀ SCIENTIFICA E DIDATTICA DI BRUNO SCARPA

FORMAZIONE SCOLASTICA E UNIVERSITARIA

- 1983—1988 Liceo Scientifico “Francesco Severi”, Venezia. Voto di diploma: 50/60
- 1988—1993 Università di Padova, Facoltà di Scienze Statistiche, Demografiche ed Attuariali. Laurea in Scienze Statistiche e Demografiche conseguita il 19 marzo 1993 con la votazione di 110/110.
Titolo della tesi: *Confronto della distribuzione del tempo fino all'allarme di alcuni schemi di controllo.*
- 1993 Esame di Stato nelle Discipline Statistiche
- 1993—1997 Università di Padova, Dipartimento di Scienze Statistiche.
Dottorato di ricerca in Statistica
Titolo della tesi: *Modelli per la concentrazione di inquinanti atmosferici combinando i dati di centraline fisse e mobili.*
Data di conseguimento del titolo: 8 luglio 1997.

POSIZIONE ATTUALE

Dal 1 settembre 2002 è ricercatore per il settore scientifico-disciplinare SECS-S/01 (Statistica) presso la Facoltà di Scienze Politiche dell'Università di Pavia.

PRECEDENTI INCARICHI

- Mar. 1993—Ott. 1993 COSES - Consorzio per lo Sviluppo Economico e Sociale Venezia della Provincia di Venezia. Collaborazione per l'analisi economica, demografica e sociale dei dati del censimento della popolazione 1991 riguardanti la Provincia di Venezia.
- Lug. 1993—Lug. 1994 OSPEDALE CIVILE DI VENEZIA Venezia
Servizio civile come obiettore di coscienza con il compito di effettuare analisi statistiche di dati clinici ed epidemiologici in collaborazione con medici e biologi.
- Feb. 1997—Lug. 1998 RAS-Riunione Adriatica di Sicurtà-Attuariato Rami Danni Milano
Analista statistico

Lug. 1998—Apr. 2000 OMNITEL s.p.a.— *Direzione Marketing – Analisi e Ricerche*
Corsico - MI Coordinatore gruppo analisi per la previsione del *churn*

Apr. 2000—Ago. 2001 TIN.IT s.p.a. (gruppo TELECOM Italia) – *Direzione Marketing*
Roma Responsabile Gruppo CRM&*Profiling*

Set. 2001—Set. 2002 H3G s.p.a. - Direzione CRM
Milano *Customer Intelligence Manager*
Azienda in *start-up*.

ATTIVITÀ DIDATTICA

A.A. 1996—1997 Facoltà di Economia, Università di Padova, ciclo di esercitazioni
tenute per il corso di Statistica (corso di base) del
Prof. C. Provasi.

A.A. 1997—1998 Facoltà di Economia, Università Cattolica del Sacro Cuore di
A.A. 1998—1999 Milano, ciclo di esercitazioni tenute per il corso di Controllo
Statistico della Qualità del Prof. U. Magagnoli.

A.A. 2002—2003 Facoltà di Scienze Politiche, Università di Pavia, affidamento del
A.A. 2003—2004 corso di Statistica Economica (60 ore) per il corso di laurea
quadriennale in Scienze Politiche.

A.A. 2002—2003 Facoltà di Scienze Politiche, Università di Pavia, affidamento del
A.A. 2003—2004 corso di Statistica Economica (40 ore) per il corso di laurea
A.A. 2004—2005 triennale in Scienze Sociali per la Cooperazione e lo Sviluppo.

A.A. 2002—2003 Facoltà di Scienze Politiche, Università di Pavia, affidamento del
corso serale di Statistica Economica (20 ore) per il corso di laurea
in Scienze Politiche.

A.A. 2002—2003 Facoltà di Scienze Politiche, Università di Pavia, ciclo di lezioni
nel corso di Statistica per i corsi di laurea triennale in Scienze
Politiche, Governo e Amministrazione e Scienze Sociali per la
Cooperazione e lo Sviluppo.

A.A. 2002—2003 Facoltà di Scienze Statistiche, Università di Padova, ciclo di
A.A. 2003—2004 lezioni per il corso di Analisi dei dati (Data Mining)
A.A. 2004—2005

A.A. 2003—2004 Facoltà di Scienze Politiche, Università di Pavia, affidamento
A.A. 2004—2005 del corso di Analisi di dati aziendali (80 ore), per il Master in
Ricerca su Società, Mercato e Territorio.

- A.A. 2003—2004 Facoltà di Sociologia, Università di Milano Bicocca, affidamento del corso di Statistica per la Ricerca Sociale - Modelli interpretativi I, per il corso di laurea quadriennale in Sociologia.
- A.A. 2004—2005 Facoltà di Scienze Statistiche, Università di Padova, ciclo di lezioni per il corso di Metodi Statistici per le Applicazioni Aziendali

SOGGIORNI STUDIO E DI RICERCA

- Dic. 1994—Set. 1995 Department of Statistics, Stanford University, CA, USA, collaborazione con il Prof. P. Switzer.
- Lug. 1996—Aug.1996 Department of Mathematics and Statistics, Lancaster University, Inghilterra, collaborazione con i Proff. P. Diggle e J. Tawn.
- Aug. 2003—Set. 2003 Biostatistical Branch, NIEHS - National Institute of Health, Research Triangle Park, NC, USA, collaborazione con il Prof. D. Dunson.
- Mag. 2004 Biostatistical Branch, NIEHS - National Institute of Health, Research Triangle Park, NC, USA, collaborazione con il Prof. D. Dunson.

PRESENTAZIONI A CONVEGNI

- Gen. 1996 *L'Informazione statistica per il Governo dell'Ambiente*, Roma.
- Apr. 1996 *Riunione scientifica dei dottorandi della XXXVIII Riunione Scientifica della SIS*, Rimini.
- Apr. 1998 *XXXIX Riunione Scientifica della SIS*, Sessione Specializzata, Sorrento.
- Apr. 1998 *XXXIX Riunione Scientifica della SIS*, Sessione Spontanea, Sorrento.
- Set. 1999 *Datawarehousing & Datamining in Telecommunications (IBC Global Conferences)*, Invited Speaker, Londra, Regno Unito.
- Giu. 2001 *I Sistemi di Business Intelligence (IIR Convegno dell'Institute of International Research)*, Relazione Invitata, Milano.
- Lug. 2002 *International Conference on Current Advances and Trends in Nonparametric Statistics*, Invited Speaker, Creta, Grecia.
- Gen. 2003 *Inaugurazione XVIII ciclo del dottorato di Padova in ricordo di Antonio C. Capelo*, Padova.

- Mag. 2004 *Expanding Methodologies for Capturing Day-Specific Probabilities of Conception*, Invited Speaker, Rockville, MD, USA.
- Giu. 2004 *Giornata Scientifica del Dipartimento di Scienze Statistiche dell'Università di Padova*.
- Giu. 2004 *Statistics in Health Sciences*, Invited Speaker, Nantes, Francia.
- Gen. 2005 *Second IMS-ISBA Joint Meeting - MCMSki*, Poster Session, Bormio (SO).
- Feb. 2005 *Giornate di Studio sulla Popolazione*, Padova.
- Apr. 2005 *Seminario di Studi della Confederazione italiana dei centri per la regolazione naturale della fertilità*, Relazione Invitata, Verona.

SEMINARI

- 28 Mar. 2001 *Data Mining nel Marketing: l'esperienza in una società di internet*
Master in Net Economy, Università di Trieste.
- 14 Nov. 2002 *Data Mining in Azienda*
Dipartimento di Scienze Statistiche, Università di Padova.
- 25 Mar. 2003 *Customer Profiling, Segmentazione e Strategie di Marketing nelle Telecomunicazioni*, Facoltà di Scienze della Comunicazione e dell'Economia, Università di Modena e Reggio Emilia.
- 29 Mag. 2003 *Modelli di data mining per la previsione del churn*,
Dipartimento di Economia Politica e Metodi Quantitativi,
Università di Pavia.
- 19 Mag. 2004 *Cervical mucus symptom and daily fecundability*,
EPI group - NIEHS - National Institute of Health,
Research Triangle Park, NC, USA
- 27 Mag. 2004 *Un'esperienza aziendale di data mining*, Facoltà di Economia,
Università del Sannio, Benevento.
- 17 Mar. 2005 *Data mining: un'esperienza nel Marketing*, Dipartimento di Statistica,
Università Ca' Foscari, Venezia.

COORDINAMENTO DI PROGETTI DI RICERCA

- Lug. 1998 — Apr. 2000 Responsabile del Gruppo di ricerca "Analisi e previsione del

Churn” formato da esperti statistici, informatici e di marketing all’interno di Omnitel s.p.a.

- Mag. 2000 — Ago. 2001 Responsabile del Progetto di ricerca “Segmentazione e Profiling degli utenti Internet” all’interno di tin.it, cui partecipavano analisti e statistici di marketing
- Set. 2001 — Set. 2002 Responsabile del Progetto “DWH e Data mining per la telefonia di terza generazione. Requisiti utente” all’interno della Divisione CRM di H3G s.p.a., cui partecipavano informatici, analisti e statistici di marketing
- Gen. 2004 — Dic. 2005 Partecipa al Progetto PRIN “Reti bayesiane causali: distribuzioni iniziali compatibili, identificabilità, aspetti dinamici, applicazioni al marketing ed al web mining.” coordinato dal prof. G. Consonni
- Gen. 2005 — Dic. 2005 Responsabile del Progetto “Biometria del ciclo mestruale” finanziato dall’Università di Pavia

INTERESSI DI RICERCA

Metodi statistici per l’analisi di problematiche aziendali

L’innovazione tecnologica permette oggi alle aziende di disporre di masse rilevanti di elementi di informazione che possono venire utilizzati per indirizzare le scelte strategiche e operative dell’azienda. In particolare molte aziende (ad esempio banche, assicurazioni, aziende di telecomunicazioni) hanno a disposizione informazioni dettagliate sui singoli clienti e sul loro utilizzo dei prodotti e servizi offerti. La mia attività di ricerca negli anni in cui ho prestato servizio nel settore privato si è concentrata principalmente sull’analisi statistica dei dati sui clienti per indirizzare le scelte strategiche in una prima fase e sul coordinamento di gruppi di ricerca su questi temi in una seconda fase.

Tra le varie attività di analisi di dati quantitativi particolarmente rilevanti sono stati: (i) l’analisi dei comportamenti e la relativa modellazione statistica della *Customer Base* mirate all’anticipazione di bisogni di utilizzo dei servizi di telefonia cellulare (in Omnitel). (ii) La definizione di indicatori di propensione al *churn*, e la predisposizione di modelli statistici per la loro previsione e individuazione dei motivi che portano alla disattivazione. (iii) Il *Profiling* e la segmentazione della *Customer Base* e dei visitatori dei siti/portali per ottenere gruppi di utenti omogenei (in tin.it). (iv) Individuazione dei profili principali dei clienti rispetto all’utilizzo dei servizi di internet. (v) Determinazione di azioni di marketing che l’azienda può effettuare per aumentare la fedeltà dei clienti e di azioni di *up-sell* e *cross-sell*. (vi) Supporto quantitativo agli altri settori di marketing per le previsioni (per *economics*, *budget*, *Business Plan* e altro). (vii) (in RAS) Analisi del rischio assicurativo nei rami danni.

Tale attività, nonostante le difficoltà a rendere pubblici i risultati di ricerca, mi ha permesso di produrre i lavori 19 e 20 per quanto riguarda le analisi di marketing e i lavori 28 e 29 nell'ambito della stima del rischio assicurativo. Il lavoro 21 è una rilettura a posteriori e sistematica con qualche esempio reale concreto dell'attività di uno statistico nel marketing di una azienda di telecomunicazioni. I lavori 33 e 34 descrivono alcune particolari specificazioni e applicazioni di modelli statistici per rispondere a problemi originati in contesti aziendali per i quali è necessario l'utilizzo di strumenti inferenziali.

Data mining e statistica

Nel mondo scientifico è ormai usuale avere a disposizione ingenti quantità di dati sia in termini di unità statistiche sia in termini di variabili disponibili (la rilevazione automatica di molte informazioni e il loro immagazzinamento permette la costituzione di grosse basi di dati utilizzabili ai fini di ricerca) richiede allo statistico di individuare strumenti innovativi ma efficaci per il trattamento dei dati. Un problema particolare affrontato riguarda la classificazione in classi rare dal quale sono scaturiti i lavori 22 e 9. Il lavoro 17 invece propone una modifica dei Modelli Additivi Generalizzati che tiene conto della presenza di autocorrelazione tra le osservazioni, problematica spesso presente in questo tipo di dati. I lavori 31 e 32 affrontano un paio di problemi concreti nati in contesti aziendali in cui è necessario analizzare ingenti quantità di dati e per i quali è necessario utilizzare specifici strumenti di analisi e di *data mining*.

L'ampia esperienza di analisi di grandi basi di dati attraverso strumenti di *data mining* mi spinto a produrre la monografia 3 in cui si presentano a partire da problemi reali i principali strumenti di analisi e data mining letti e interpretati dal punto di vista dello statistico.

Modelli statistici per l'analisi di indicatori biometrici del ciclo mestruale

Il ciclo mestruale è una delle principali funzioni biologiche che regola la riproduzione umana; l'utilizzo di tecnologie sempre più avanzate in campo medico ha permesso, negli ultimi cinquant'anni di studiare e di capire sempre meglio i complessi sistemi biologici del ciclo riproduttivo. L'analisi delle grandezze legate a tale ciclo ha poi molto spesso dei risvolti pratici nel comportamento delle persone per quanto riguarda l'attività sessuale e la vita riproduttiva. Le problematiche di modellazione statistica sono molto complesse in particolare considerando le informazioni rilevabili attraverso strumenti non invasivi e relativamente economici. Nell'ambito di ricerche in questo campo ho affrontato alcune tematiche in particolare legate alla stima della probabilità di concepimento in un ciclo quando si hanno a disposizione informazioni ridotte. Il lavoro 12 affronta tale problema attraverso una modellazione frequentista, mentre nei lavori 11, 13, e 23 il tema è affrontato con un approccio bayesiano. I lavori 14, 15 e 30 affrontano attraverso un approccio decisionale bayesiano il problema di identificazione da parte delle coppie di una decisione che massimizzi la loro utilità nella scelta dei giorni in cui avere rapporti sessuali o dei comportamenti da assumere in termini di variabili epidemiologiche (fumo, alcool, caffè, ecc.). Il lavoro 6 analizza un problema analogo di identificazione della finestra fertile a partire da poche informazioni attraverso un approccio classico.

Analisi spazio-temporale di dati ambientali

La concentrazione di sostanze chimiche presenti nell'atmosfera costituisce uno degli indicatori principali per la segnalazione di situazioni di inquinamento. L'analisi di tali dati è quindi un tema rilevante per il controllo dell'atmosfera. Nella tesi di dottorato, 2, mi sono concentrato principalmente nella identificazione e nella metodologia di stima di alcuni modelli per la previsione della concentrazione di sostanze inquinanti. A partire da tale lavoro ho trattato alcuni aspetti particolari dell'analisi come l'identificazione di mappe spaziali nel lavoro 7, l'utilizzo di centraline mobili e una modellazione completamente stocastica nel lavoro 8, il posizionamento di nuove centraline attraverso un approccio non parametrico nei lavori 18 e 27 e l'effetto della presenza di dati mancanti utilizzando un approccio di stima non parametrico nel lavoro 8.

Applicazioni Biomediche

Le problematiche di carattere epidemiologico e medico sono un campo dove l'applicazione di modelli statistici ai dati osservati è necessaria per comprendere e interpretare i fenomeni. Il lavoro 5 è una analisi sull'effetto dell'utilizzo della classificazione delle malattie attraverso i DRG in un ospedale e il lavoro 24 consiste in un'analisi sul rischio di nascite con malformazioni congenite, analizzato attraverso modelli non parametrici.

Applicazioni Didattiche

Il lavoro 4 ha scopo prevalentemente didattico. Intende essere un testo divulgativo per la comprensione da parte di studenti, con alle spalle un solo esame di statistica di base, dei concetti fondamentali del modello lineare e delle sue più semplici modificazioni. Nel lavoro 16 vengono presentati a scopo divulgativo, indirizzato a ostetriche e operatori di consultori familiari, i principali risultati scientifici sulla dimensione e posizionamento della finestra fertile nel ciclo riproduttivo femminile. Il lavoro 10 è una presentazione, a scopo illustrativo, dell'utilizzo di modelli statistici complessi per la valutazione di programmi sociali quando la modellazione più semplice non è sufficiente a cogliere la complessità della realtà.

PUBBLICAZIONI

MONOGRAFIE

1. Scarpa, B. (1993). *Confronto della distribuzione del tempo fino all'allarme di alcuni schemi di controllo*. Tesi di laurea, Università di Padova.
2. Scarpa, B. (1997). *Modelli per la concentrazione di inquinanti atmosferici combinando i dati di centraline fisse e mobili*. Tesi di dottorato, Università di Padova.
3. Azzalini A., Scarpa, B. (2004). *Analisi dei dati e data mining*. Springer, Milano, ISBN 88-470-0272-9.
4. Scarpa, B. (2005). *Il modello lineare: una presentazione a partire da dati reali*. Arvan, Mira-Venezia, ISBN 88-87801-15-0.

ARTICOLI SU RIVISTE

5. Selle V. e Scarpa B. (1995). La qualità dei dati delle dimissioni ospedaliere nell'Ospedale Civile di Venezia: aspetti metodologici e valutativi ai fini dell'utilizzo dei DRG, *Epidemiologia e prevenzione*, **19**, pp. 342—348.
6. Colombo B. e Scarpa B. (1996). Calendar methods of fertility regulation: a rule of thumb, *Statistica*, **LVI**, pp. 3—14.
7. Capelo A. C., Masarotto G. e Scarpa B. (1996). Costruzione di mappe delle concentrazioni di inquinanti atmosferici combinando i dati di centraline fisse e mobili, *Annali di Statistica*, Serie **X**, vol.**10**, pp. 257—265.
8. Scarpa B. (2001). La concentrazione di SO₂ combinando i dati raccolti da centraline fisse e mobili: un modello state space, *Statistica*, **LXI**, pp. 707—730.
9. Scarpa B., Torelli N. (2004) Selecting the training set in classification problems with rare events. In: M. Vichi, P. Monari, S. Mignani, A. Montanari, (curatori) *New Developments in Classification and Data Analysis*, Springer-Verlag, pp. 39—46.
10. Scarpa B. (2004) Networks, Uncertainty and Stochastic models. In: E. Tezza, *Evaluating social programmes. The relevance of relationships*, ILO, Torino, pp. 57—61.
11. Scarpa B., Dunson D.B., Colombo B. (2005) Cervical mucus secretions on the day of intercourse: An accurate marker of highly fertile days, *Journal of Obstetrics & Gynecology and Reproductive Biology*, sottoposto a revisione, accettato con minori revisioni.
12. Colombo B., Mion A., Passarin K., Scarpa B. (2005) Cervical mucus symptom and daily fecundability: first results from a new data base. *Statistical Methods in Medical Research*, sottoposto a revisione.
13. Scarpa B., Dunson D.B., (2005) Bayesian selection of predictors of day-specific conception probabilities, *Paediatric and Perinatal Epidemiology*, sottoposto a revisione.
14. Scarpa B., Dunson D.B., Colombo B. (2005) Bayesian Selection of Optimal Rules for Timing Intercourse Using Calendar and Mucus, *Contraception*, sottoposto a revisione.
15. Scarpa B., Dunson D.B., (2005) Making informed decisions about behavior for couples attempting pregnancy, *Statistical Methods in Medical Research*, sottoposto a revisione.
16. Scarpa B. (2005). Non-parametric space-time modeling of SO₂ in presence of many missing data, *Statistical Methods and Applications-JISS*, in corso di stampa.

PUBBLICAZIONI IN ATTI DI CONVEGNI

17. Scarpa B. (1998). Modelli Additivi Generalizzati con autocorrelazione tra le osservazioni. In *Atti della XXXIX Riunione Scientifica della Società Italiana di Statistica*, Sessioni spontanee (Sorrento, 15–18 aprile 1998), pp. 857—864.

18. Masarotto, G, Scarpa B. (1998). Analisi spazio-temporale di inquinanti atmosferici e posizionamento di centraline: un approccio non-parametrico, In *Atti della XXXIX Riunione Scientifica della Società Italiana di Statistica*, Sessioni plenarie e specializzate (Sorrento, 15–18 aprile 1998), pp. 195—206.
19. Scarpa B. (1999). Customer Profiling, Segmentation and Marketing Strategies. In *Datawarehousing & Datamining in Telecommunications (IBC Global Conferences)* (Londra, 14–15 settembre 1999).
20. Scarpa B. (2001). Fidelizzare i clienti: come passare dai dati alle azioni. In *I sistemi di Business Intelligence (Convegno IIR)* (Milano, 26–27 giugno 2001).
21. Scarpa B. (2003). Il ruolo dello statistico di marketing nelle aziende di telecomunicazioni. In *Atti della Giornata di Inaugurazione XVIII ciclo del dottorato di Padova in ricordo di Antonio C. Capelo* (Padova, 9 gennaio 2003), pp. 97—106.
22. Scarpa B., Torelli N. (2003) Selecting the training set in classification problems with rare events. In *Atti del convegno CLADAG 2003* (Bologna, settembre 2003).
23. Scarpa B. (2004) Cervical mucus symptom and daily fecundability. In *Proceedings of the International Conference on Statistics in Health Sciences* (Nantes, June 23–25, 2004)
24. Rivellini G., Rosina A., Scarpa B. (2004) Late maternal age and risk of congenital malformation, In *Atti della XLII Riunione Scientifica della Società Italiana di Statistica*, Sessioni spontanee (Bari, 9–11 Giugno 2004), pp.19—22.
25. Bortot P., Scarpa B., Masarotto G. (2005) Previsioni sequenziali delle lunghezze del ciclo mestruale, In *Atti delle Giornate di Studio sulla Popolazione*, (Padova, 16–18 Gennaio 2005).
26. Scarpa B. (2005) La finestra fertile “statistica”, In *Atti del Seminario di Studi della Confederazione Italiana dei Centri per la Regolazione Naturale della Fertilità*, (Verona, 8–9 Aprile 2005).

RAPPORTI TECNICI E MANOSCRITTI

27. Arbia G., Lafratta G., Scarpa B. (1996), Evaluating and updating the sample design for the concentration of SO₂ in Padua, Università degli Studi G. d’Annunzio di Pescara, Dipartimento di Metodi Quantitativi e Teoria Economica, *Quaderni di statistica*, n. 4/1996.
28. Scarpa B. (2001), Il grado di invalidità permanente nella polizza infortuni: una valutazione di variabili predittive mediante modello a rischi proporzionali di Cox, Università degli Studi di Padova, Dipartimento di Scienze Statistiche, *Working Paper*, n. 2001.6.
29. Scarpa B. (2001), La tariffa incendio e furto per le auto: un modello di Cox, Università degli Studi di Padova, Dipartimento di Scienze Statistiche, *Working Paper*, n. 2001.7.

30. Scarpa B. (2004), Choice of an "optimal rule" for avoid conception using calendar and mucus data, Università degli Studi di Pavia, Dipartimento di Statistica ed Economia Applicate, *Working Paper*, n.19.
31. Scarpa B. (2004), La previsione del *churn*: problemi e possibili soluzioni di *data mining* nella realtà della telefonia cellulare, Università degli Studi di Pavia, Dipartimento di Statistica ed Economia Applicate, *Working Paper*, n.20.
32. Scarpa B. (2004), Confronto tra modelli di *data mining* per la previsione della domanda di energia elettrica, Università degli Studi di Pavia, Dipartimento di Statistica ed Economia Applicate, *Working Paper*, n.21.
33. Scarpa B. (2004), La *Customer statisfaction* per un'azienda di servizi informatici. Impostazione e analisi di un'indagine via web, Università degli Studi di Pavia, Dipartimento di Statistica ed Economia Applicate, *Working Paper*, n.22.
34. Scarpa B. (2005), Lo stress in azienda. Modelli di analisi di un'indagine per l'identificazione delle cause di stress, Università degli Studi di Pavia, Dipartimento di Statistica ed Economia Applicate, *Working Paper*, n.23.

Il sottoscritto

COGNOME: Scarpa
NOME: Bruno
CODICE FISCALE: SCRBRN69A29L736I
NATO A: Venezia
IL: 29/01/1969 SESSO: M

consapevole che, secondo quanto previsto dagli artt. 46 e 47 D.P.R. n. 445/00, le dichiarazioni mendaci sono punite ai sensi del codice penale e delle leggi speciali in materia,

DICHIARA:

che quanto indicato nel presente curriculum corrisponde a verità.

Venezia, 8 giugno 2005

Bruno Scarpa