



**STUDI IN RICORDO
DI
CARLO CECCHI**

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO
2015

Tutti i diritti di traduzione, riproduzione e adattamento,
totale o parziale, con qualsiasi mezzo (comprese
le copie fotostatiche e i microfilm) sono riservati

© Copyright 2015 by Università degli Studi di Bari Aldo Moro
www.uniba.it

Prima edizione: dicembre 2015

ISBN 978-88-88793-67-2

Editing: F. D. d'Ovidio

Sommario

Presentazione	pag.	7
Giovanni Girone, Antonella Massari		
La differenza media della variabile F di Snedecor	«	9
Nicola Bartolomeo, Monica Carbonara, Gabriella Serio, Antonella Mincuzzi, Aldo Minerba, Paolo Trerotoli		
Applicazione di un modello bayesiano per la mappatura del rischio relativo in un'area ad elevato rischio ambientale	«	15
Massimo Bilancia, Salvatore Bello, Michele Scalera, Domenico Viola		
A personalized recommender system for a luxury e-commerce website. An approach based on market basket data	«	25
Giovanni Girone, Claudia Marin		
L'effetto dei valori negativi sulla differenza media e sul rapporto di concentrazione	«	41
Dante Mazzitelli, Najada Firza		
Indicatori sanitari di misurazione e valutazione delle performance	«	49
Giuseppina Sacco, Pietro Sacco		
La valenza demografica delle componenti naturali e sociali nelle previsioni della popolazione barese	«	63
Francesco D. d'Ovidio, Rossana Mancarella, Leonardo Mariella, Ernesto Toma, Angelo Valenzano		
Determinanti socio-esperienziali delle performance finanziarie	«	79
Domenico Summo, Tommaso Pepe		
Misura del benessere e performance di un territorio	«	95
Mauro Bisceglia		
Analisi dei paradigmi latenti della employee satisfaction in ambiente bancario tramite un modello SEM	«	111
Lucia Mongelli		
La distribuzione di Meixner per la rappresentazione degli indici elementari	«	123
Caterina Marini, Vittorio Nicolardi, Giovanni Vannella		
Rapporti di lavoro non standard: impiego, utilità e criticità	«	131

Giovanni Girone, Fabio Manca, Angela Maria D'Ugento <i>An implementation of the Gini concentration ratio decomposition by sources</i>	« 153
Rossana Mancarella, Nunzia Ribecco, Ernesto Toma <i>Dimensioni latenti dell'attitudine al gioco di fortuna in età adolescenziale</i>	« 161
Angela Maria D'Ugento, Vito Ricci, Ernesto Toma <i>A proposal for the performance evaluation in research of the top 25 Countries in the world</i>	« 177
Agata Maria Madia Carucci, Antonello Costanzo, Giovanni Vannella <i>Studi di settore tra lotta all'evasione e Statistica</i>	« 191
Francesco Campobasso, Annarita Fanizzi <i>A regression model for new assignments to fuzzy groups</i>	« 213
Laura Antonucci, Marina Basile, Corrado Crocetta, Viviana D'Addosio, Francesco D. d'Ovidio <i>Fidelizzazione degli utenti del sito web dell'Università di Bari</i>	« 227
Giovanni Girone, Antonella Nannavecchia <i>Modelli distributivi per l'analisi delle dinamiche dei mercati finanziari</i>	« 245

Presentazione

Poco più di dieci anni fa, il 26 settembre 2005, per l'improvviso acuirsi di una patologia di cui era già sofferente, all'età di quasi 67 anni si spegneva Carlo Cecchi.

Nato a Bari il 6 novembre del 1938, Carlo Cecchi è stato professore ordinario di Statistica, Presidente del Consiglio del Corso di laurea in Economia e Commercio dell'Università degli Studi di Bari, nonché Direttore dell'Istituto di Statistica e poi del Dipartimento di Scienze Statistiche della medesima università. Fu tra i promotori della sede universitaria di Foggia, di cui è stato anche il primo coordinatore del Corso di Diploma in Economia e Gestione dei Servizi Turistici. Contribuì all'istituzione del Corso di laurea in Economia aziendale a Taranto, che coordinò dal 1996 al 2000, anno in cui venne eletto Preside della Facoltà di Economia di Bari, carica che gli fu poi rinnovata e che svolse fino all'ultimo con profonda abnegazione. Il suo impegno nella sede barese non gli impedì mai di avere una particolare attenzione verso la crescita del polo universitario tarantino, in cui si sviluppò, infatti, la II Facoltà di Economia dell'Università degli Studi di Bari.

Dottore commercialista (dal 1965) e iscritto, fin dall'istituzione (1995), nel Registro dei Revisori contabili presso il Ministero di Grazia e Giustizia, fu anche componente del Consiglio dell'Ordine dei Dottori Commercialisti di Bari. Membro eletto dell'Istituto Internazionale di Statistica e socio di più associazioni scientifiche italiane ed internazionali, fece anche parte del Consiglio Direttivo della Società Italiana di Statistica (1994-98). Fu componente del *Comitato tecnico-scientifico per la programmazione* della Regione Puglia (1980-82), del *Comitato scientifico per l'attuazione della L. 142/90* sempre per la Regione Puglia (1991-92), e membro effettivo della *Commissione di garanzia elettorale* della Puglia dal 1994 al 2002. Dal 1986 fu tesoriere dell'Accademia Pugliese delle Scienze, e dal 1996 presidente dell'IPRES - Istituto Pugliese di Ricerche Economiche e Sociali.

Autore di apprezzati saggi didattici e di numerosi contributi scientifici pubblicati su riviste specializzate o presentati in convegni nazionali ed internazionali, nel 1995 curò anche il rapporto sullo stato degli uffici comunali di Statistica della Provincia di Bari nel quadro dell'indagine conoscitiva avviata dalla Commissione di Garanzia per l'Informazione Statistica.

A distanza di tanti anni, la sua opera e la sua profonda umanità non sono state dimenticate e, a testimonianza di ciò, gli amici e colleghi che con lui hanno costituito il Dipartimento di Scienze statistiche (poi denominato a suo nome) hanno voluto dedicargli la presente raccolta di scritti di argomento statistico e gestionale.

*Gli amici e colleghi
del Dipartimento di Scienze statistiche*

Fidelizzazione degli utenti del sito web dell'Università di Bari*

Laura Antonucci², Marina Basile¹, Corrado Crocetta²,
Viviana D'Addosio³, Francesco D. d'Ovidio⁴

¹ *Istituto tecnico-commerciale e linguistico "Marco Polo", Bari*

² *Dipartimento di Economia, Università degli Studi di Foggia*

³ *IPSIA "Archimede", Andria*

⁴ *Dipartim. di Scienze Economiche e Metodi Matematici, Univ. degli Studi di Bari Aldo Moro*

Riassunto: I siti web del settore formativo sono stati studiati da molte prospettive diverse, tecniche, gestionali, ecc.; in particolare, nel 2001 Zhang e Dran hanno sviluppato un quadro teorico per valutare la qualità del sito web dal punto di vista della soddisfazione degli utenti. Seguendo tale impostazione, in questo lavoro si valuta la qualità percepita dagli utenti del sito web dell'Università degli Studi di Bari utilizzando mappe visive e analisi fattoriale per individuare le dimensioni primarie dei giudizi degli utenti. Le variabili latenti derivanti da questa analisi preliminare sono state poi utilizzate per studiare i profili-utente e per valutare le dimensioni più importanti per la fidelizzazione degli utenti, utilizzando come *proxy* la frequenza di accesso al sito.

Keywords: Website universitari, Analisi Fattoriale, Analisi delle Corrispondenze Multiple, Mappe visuali, Fidelizzazione, Profili-utente.

1. Premessa

I siti web delle università stanno diventando componenti indispensabili della formazione, non solo per le Università telematiche. Per via delle normative si sono succedute nel corso degli anni, infatti, anche in una università tradizionale la maggior parte delle operazioni relative ai corsi (ad esempio, informazioni di carattere

* Il presente articolo è frutto del lavoro comune degli Autori, ma è attribuibile a L. Antonucci il § 5. M. Basile il § 3, a C. Crocetta i § 1 e 6, a V. D'Addosio il § 2 e a F. D. d'Ovidio il § 4.

generale, il contatto tra insegnanti e studenti, la prenotazione degli esami di profitto, e ormai anche la verbalizzazione) avviene on-line. La qualità e la fruibilità dei siti web sono quindi sempre più importanti, e si deve porre grande attenzione per migliorare queste caratteristiche.

Tuttavia, un sito web è composto di molti aspetti, che risulta a volte difficile ottimizzare contemporaneamente. Valutando la qualità del sito web dal punto di vista della soddisfazione degli utenti, è tuttavia possibile sviluppare un quadro teorico affidabile, come già fatto in passato da Zhang e Dran (2001).

In questo lavoro, si vuole dunque analizzare quali possono essere gli aspetti che maggiormente impattano sulla percezione di qualità degli utenti e anche sulla loro assiduità. A tale scopo, i dati relativi alla qualità percepita dagli utenti del sito web dell'Università degli Studi di Bari (misurati con scala di valutazione a 5 livelli) vengono prima analizzati usando analisi fattoriale e mappe visive (analisi delle corrispondenze) per individuare le dimensioni principali delle valutazioni fornite; le variabili latenti derivanti da questa analisi, espresse in punteggi standardizzati, vengono poi utilizzati sia per la progettazione di profili-utente non influenzati da fattori di disturbo (rumore bianco, errori di rilevazione, ecc.) e per valutare quale dimensione latente è più importante per gli utenti in termini di fidelizzazione alla sito, interpretando come *proxy* di tale aspetto (*coeteris paribus*) la frequenza di accesso al sito.

2. Analisi esplorativa degli aspetti rilevati

Va detto, per prima cosa, che la variabile prescelta per rappresentare la fidelizzazione degli utenti (appunto la frequenza di accessi al sito) presenta una distribuzione alquanto critica a fini statistici. Infatti, su 1.049 rispondenti (41% di genere maschile e 59% di genere femminile), 112 accedono al sito più volte al giorno, 317 più volte a settimana, 390 circa una volta a settimana, 223 circa una volta al mese, e soltanto 7 non vi accedono mai. Per la ridotta numerosità di quest'ultima modalità, l'analisi statistica ha dunque richiesto di aggregarla a quella precedente.

Per inquadrare il fenomeno, è stata poi effettuata una analisi bivariata della *proxy* "accesso al sito", così modificata, con tutte le altre variabili presenti nel database, sia in relazione alle caratteristiche dei rispondenti, sia rispetto ai giudizi da questi espressi sui vari aspetti del sito.

Innanzitutto, dalla Tabella 1 si evince la distribuzione percentuale degli accessi ai siti web secondo il tipo di Corso di studio frequentato e alla Facoltà (ove le "Facoltà" qui rappresentano aggregazioni di dipartimenti all'incirca omogenei, grossomodo corrispondenti alle Facoltà ante L. 240/2010, la c.d. "Legge Gelmini").

Tabella 1. Distribuzione percentuale degli utenti secondo il Corso di studio e la Facoltà, per frequenza di accesso al sito.

Posizione accademica	Frequenza d'accesso al sito web				Totale
	Mai / occasionalmente	Circa una volta a settimana	Più volte a settimana	Più volte al giorno	
CORSO DI STUDIO					
Corso di laurea triennale	50,9	57,7	58,1	69,6	57,6
Corso di laurea magistrale	20,9	17,2	21,1	19,7	19,4
Corso di laurea magistrale a ciclo unico	26,5	24,1	18,6	9,8	21,4
Corsi universitari di formazione finalizzata	1,3	0,8	0,6	0,0	0,8
Dottorato di ricerca	0,4	0,2	1,6	0,9	0,8
FACOLTÀ					
Agraria e Veterinaria	6,7	6,9	9,5	8,0	7,8
Scienze chimico-ambientali	11,6	11,2	8,6	2,7	9,6
Scienze matematiche e informatiche	20,9	11,5	9,5	8,0	12,6
Medicina	11,6	8,8	6,7	4,5	8,3
Lingue, lettere, storia e filosofia	5,3	13,3	15,9	25,9	13,7
Scienze umane	17,3	15,7	13,7	8,0	14,6
Giurisprudenza	13,3	18,4	15,0	8,9	15,2
Scienze politiche e Giuridico-econ. Jonico	3,1	3,5	4,8	4,5	3,9
Economia	10,2	10,7	16,3	29,5	14,3
ANNO DI CORSO					
Primo	17,4	20,8	27,8	25,0	22,6
Secondo	17,8	16,4	14,2	23,2	16,8
Terzo	16,1	17,4	20,5	12,5	17,6
Quarto (o primo di laurea magistrale)	14,4	10,5	11,0	16,1	12,1
Quinto (o secondo di laurea magistrale)	10,4	9,8	11,0	6,2	9,9
Sesto	1,3	1,5	1,3	0,0	1,2
Fuori corso	22,6	23,6	14,2	17,0	19,8
Totale	100,0	100,0	100,0	100,0	100,0

Dalla tabella emerge che fra gli studenti dei corsi di laurea triennali è significativamente maggiore ($p < 0,01$)¹ la frequenza di quelli che accedono al sito web “Più volte al giorno”, rispetto, ad esempio, agli studenti dei corsi di laurea magistrale a

¹ Verifiche statistiche effettuate con il test del rapporto di massima verosimiglianza; per le particolarità e i limiti applicativi di questo test, così come del successivo test ², si veda, ad es. Delvecchio 2015. Trattandosi di interazioni, la scelta di descriverne l'effetto in termini di distribuzione delle variabili di contrasto entro ogni frequenza di accesso o viceversa è legata unicamente a motivi di rappresentatività delle percentuali rilevate: invero, la distribuzione degli utenti per frequenza di accesso al sito web è nota, e numericamente rilevante per ogni numerosità, laddove alcune modalità delle variabili di contrasto fanno rilevare un numero di rispondenti alquanto esiguo.

ciclo unico, che hanno dichiarato di accedervi solo occasionalmente o addirittura di non accedervi mai in percentuale maggiore di altri.

Anche l'iscrizione a corsi di Dipartimenti afferenti a differenti Facoltà fa registrare una frequenza di accessi al sito web dell'Università di Bari significativamente diversificata ($p < 0,001$): gli alunni che in maggior percentuale hanno dichiarato di accedervi più volte al giorno sono infatti quelli di Economia e di Lingue, lettere, storia e filosofia, al contrario degli iscritti a Scienze matematiche e informatiche e a Scienze umane, fra i quali, invece, si registra una maggiore incidenza percentuale di coloro che accedono poco o nulla al sito medesimo.

L'iscrizione a corsi di Dipartimenti afferenti a differenti "Facoltà" fa registrare una frequenza di accessi al sito web dell'Università di Bari significativamente diversificata ($p < 0,001$): gli alunni che in maggior percentuale hanno dichiarato di accedervi più volte al giorno sono infatti quelli di Economia e di Lingue, lettere, storia e filosofia, al contrario degli iscritti a Scienze matematiche e informatiche e a Scienze umane, fra i quali, invece, si registra una maggiore incidenza percentuale di coloro che accedono poco o nulla al sito medesimo.

Anche l'anzianità di iscrizione ai corsi universitari è parimenti connessa con la frequenza di accesso al sito in modo significativo ($p < 0,02$): infatti, mentre la maggior quota di studenti con accesso nullo oppure occasionale è composta da fuori corso, coloro che accedono al sito con maggior frequenza sono in maggior percentuale studenti del primo e secondo anno delle lauree triennali.

Non si rilevano, invece, relazioni statisticamente significative tra la frequenza di accesso al sito e le altre caratteristiche degli utenti (genere, residenza ecc).

Dalla distribuzione percentuale degli utenti secondo il numero di accessi, distinti invece per motivi di connessione al sito², emerge che nel 37% dei casi gli utenti accedono al sito web con una frequenza pari ad un accesso a settimana, nel 30% più volte a settimana e quasi l'11% quotidianamente, anche più volte al giorno. Accedono con una maggiore frequenza soprattutto per conoscere la variazione delle aule e gli orari delle lezioni, o per leggere news sugli eventi, mentre il numero di accessi tende a ridursi in modo significativo (fino a diventare occasionali o meno) per accedere alla segreteria on line, per prenotare gli esami on-line (almeno fino al-

² Motivi che sono molteplici, e ogni utente poteva contrassegnarne uno o molti, cosicché la verifica statistica delle eventuali influenze che essi potessero avere sulla frequenza di accesso al sito web ha richiesto la creazione di altrettante variabili dummy, da valutare singolarmente con il test ² (qui affidabile data la cospicua numerosità dei subcampioni relativi alle singole motivazioni, dei quali il minore ammonta a 300 casi). Si rileva qui un'influenza statisticamente significativa sull'accesso per tutti i motivi di connessione, quasi tutti con $p < 0,001$, fuorché per la *prenotazione on line degli esami* ($p < 0,003$) e per l'*accesso al profilo/segreteria on-line* ($p < 0,005$).

la pratica obbligatorietà della modalità telematica per la prenotazione), per scaricare modulistica oppure per ricercare informazioni sugli uffici (Tabella 2).

Tabella 2. *Distribuzione percentuale degli utenti secondo la frequenza di accesso, per motivo per cui si connettono al sito web dell'Università di Bari*

Motivi connessione	Frequenza di accesso al sito web				Totale
	Mai / occasionalmente	Circa una volta a settimana	Più volte a settimana	Più volte al giorno	
Variazione aule/orari delle lezioni	11,1	36,0	35,4	17,5	100,0
Prenotazione on line degli esami	19,3	37,0	31,6	12,1	100,0
Consultazione calendario esami	17,7	37,7	32,0	12,6	100,0
Download modulistica	18,1	36,7	32,4	12,8	100,0
Accesso al profilo/Segreteria on-line	20,3	37,0	31,7	11,0	100,0
Consultazione bacheche docenti	14,9	32,4	37,5	15,2	100,0
Ricerca informazioni su uffici ecc.	17,2	32,1	35,8	14,9	100,0
Consultazione news su eventi ecc.	11,5	33,3	37,4	17,8	100,0
Controllare la posta elettronica	16,1	36,2	33,4	14,3	100,0
Totale	21,9	37,2	30,2	10,7	100,0

La Tabella 3 riporta, infine, le interazioni significative tra frequenza di accesso al sito e giudizi espressi sui diversi aspetti del sito medesimo, evidenziando, con l'impostazione già utilizzata nella Tabella 1, come i giudizi espressi relativamente ai differenti aspetti del sito web si distribuiscono nei vari gruppi di utenti distinti in base ai loro accessi al medesimo.

Tra gli utenti che hanno dichiarato di accedere al sito più volte al giorno si rileva sempre, come è naturale, una quota maggiore di giudizi positivi, mentre coloro che invece effettuano un numero scarso o nullo di accessi in maggioranza hanno espresso un giudizio neutro (né positivo, né negativo).

Va sottolineato che rispetto alla variabile "*Livello di approfondimento e dettaglio dei contenuti*" ha espresso un giudizio neutrale la maggior parte degli utenti senza molta distinzione in base alla frequenza di accesso al sito, anche se il fenomeno appare più evidente nell'ipotesi di limitatissimo numero di accessi, mentre nel caso opposto (più volte al giorno) i giudizi positivi sono un po' più frequenti.

La situazione cambia quando si osserva la variabile "*Comprensibilità del lessico utilizzato*", relativamente alla quale si rivela una quota maggiore di giudizi addirittura "molto positivi" tra coloro che effettuano frequentissimi accessi al sito, ma i giudizi restano comunque positivi anche tra coloro che effettuano un numero inferiore di accessi, perfino occasionalmente.

Tabella 3. Distribuzione percentuale degli utenti secondo i giudizi espressi sui diversi aspetti del sito web, per frequenza di accesso al sito.

Giudizi espressi		Frequenza d'accesso al sito web				Totale
		Mai / occasionalmente	Circa una volta a settimana	Più volte a settimana	Più volte al giorno	
Livello di utilità delle informazioni pubblicate (p<0,001)	Molto Negativo	3,9	2,6	2,9	2,8	3,0
	Negativo	10,5	11,1	8,6	13,9	10,5
	Né negativo né positivo	41,7	31,1	25,7	20,3	30,7
	Positivo	33,8	38,9	40,3	42,6	38,6
	Molto Positivo	10,1	16,3	22,5	20,3	17,2
Livello di approfondimento e dettaglio dei contenuti (p<0,03)	Molto Negativo	4,8	5,1	6,9	3,7	5,5
	Negativo	17,5	19,6	16,7	15,7	17,9
	Né negativo né positivo	50,2	37,9	34,4	38,9	39,6
	Positivo	20,5	29,9	31,9	28,7	28,3
	Molto Positivo	7,0	7,5	10,1	13,0	8,7
Comprensibilità del lessico utilizzato (p<0,04)	Molto Negativo	1,8	1,3	2,2	1,8	1,7
	Negativo	6,1	6,2	6,3	11,0	6,7
	Né negativo né positivo	24,6	20,1	14,8	23,0	19,8
	Positivo	40,3	42,3	40,1	26,6	39,5
	Molto Positivo	27,2	30,1	36,6	37,6	32,3
Segnalazione errori o mal funzionamenti durante la navigazione (p<0,005)	Mai	1,9	3,2	5,2	3,4	3,5
	Quasi mai	16,3	17,7	19,2	20,5	18,1
	Quasi sempre	52,6	55,6	61,5	65,9	57,7
	Sempre	0,5	0,0	0,0	0,0	0,1
	Mai capitato	28,7	23,5	14,1	10,2	20,6
Durata interruzioni (p<0,04)	Mai	0,9	1,3	2,8	5,5	2,1
	Quasi mai	12,4	11,8	11,4	7,4	11,3
	Quasi sempre	67,1	67,0	59,2	59,6	63,9
	Sempre	19,6	19,9	26,6	27,5	22,7
Tempo di download (p<0,02)	Insufficiente	2,7	1,6	1,6	1,9	1,9
	Mediocre	5,7	7,6	6,0	6,5	6,6
	Sufficiente	31,0	25,8	17,1	19,4	23,6
	Buono	45,1	40,8	50,6	49,1	45,6
	Ottimo	15,5	24,2	24,7	23,1	22,3
Visualizzazione del sito su qualsiasi browser (p<0,05)	Insufficiente	3,1	2,6	2,6	3,6	2,8
	Mediocre	7,4	8,4	6,1	6,4	7,3
	Sufficiente	32,0	28,4	18,6	25,5	25,9
	Buono	38,6	36,1	47,1	39,1	40,3
	Ottimo	18,9	24,5	25,6	25,5	23,7
Adeguatezza trattazione dei contenuti (p<0,03)	1	1,0	1,6	3,1	7,9	2,7
	2	12,6	6,3	11,8	4,8	9,1
	3	43,7	36,5	30,4	27,0	34,9
	4	29,1	37,6	42,3	38,1	37,4
	5	13,6	18,0	12,4	22,2	15,9

Giudizi espressi		Frequenza d'accesso al sito web				Totale
		Mai / occasionalmente	Circa una volta a settimana	Più volte a settimana	Più volte al giorno	
Terminologia comprensibile e non ambigua ($p < 0,02$)	1	4,9	2,4	3,8	1,0	3,2
	2	11,5	12,3	8,3	16,8	11,4
	3	35,4	29,7	30,3	21,5	30,2
	4	35,8	37,5	36,3	35,5	36,6
	5	12,4	18,1	21,3	25,2	18,6
Identificazione utenti ($p < 0,04$)	1	2,2	1,3	1,6	1,8	1,7
	2	8,0	3,7	4,5	2,8	4,8
	3	20,5	18,7	14,7	20,2	18,0
	4	44,5	45,4	40,4	35,8	42,6
	5	24,9	30,9	38,8	39,4	32,9
Giudizio complessivo sul sito ($p < 0,02$)	1	3,0	1,3	2,8	0,9	2,1
	2	8,7	8,2	9,5	1,8	8,0
	3	37,8	35,6	32,2	38,4	35,4
	4	44,4	48,5	47,6	42,8	46,7
	5	6,1	6,4	7,9	16,1	7,8
Totale		100,0	100,0	100,0	100,0	100,0

Osservando le variabili “*Segnalazione errori o malfunzionamenti durante la navigazione*” e “*Durata interruzioni*”, si nota che sia gli utenti che visitano il sito più volte al giorno, sia quelli che non vi accedono mai o quasi mai, nella maggior parte dei casi, hanno dichiarato che tali servizi si attivano “quasi sempre”. La differenza fra i due gruppi estremi si nota per i giudizi più orrimentali.

Sia riguardo al “*Tempo di download*” che riguardo alla *Visualizzazione del sito su qualsiasi browser* e alla “*Adeguatezza della trattazione dei contenuti*”, la maggior quota di utenti si ritrova per un “buon” giudizio, ma mentre coloro che effettuano accessi frequenti o frequentissimi presentano un secondo massimo per i giudizi più ottimali, coloro che limitano al minimo il numero di accessi (mai/occasionalmente), sono sbilanciati verso un giudizio “sufficiente” (che rappresenta addirittura la moda per la seconda variabile. Risulta sempre evidente la maggior quota di valutazioni massime per chi accede più volte al giorno.

Relativamente alle variabili “*Terminologia comprensibile e non ambigua*”, “*Identificazione utenti*”, “*Giudizio complessivo sul sito*”, la frequenza di accesso al sito web influenza in modo evidente non sembra influenzare moltissimo il giudizio espresso dagli utenti nei confronti di tali aspetti, in termini di moda e mediana, che infatti si attestano intorno al valore 4. Tuttavia sono ancora significativamente differenti le valutazioni estreme (1 e 5).

3. Analisi fattoriale per definire le dimensioni di qualità

Sono state dapprima selezionate tutte le variabili di valutazione degli aspetti del sito web. Inizialmente, tramite analisi delle componenti principali categoriali³ sono state esaminate tutte le 46 variabili del questionario “utenti”, estraendo le componenti con autovalore maggiore di 1,1, ossia in grado, ciascuna, di spiegare più variabilità degli item che vi contribuiscono; eliminando poi, con procedura di eliminazione *backward stepwise*, le variabili con comunanza <0,51; si è giunti così a identificare un insieme di 26 variabili connesse a 6 componenti. Le componenti individuate spiegano, in tutto, oltre il 70% della variabilità complessiva (Tab. 1).

Peraltro, il test di sfericità di Bartlett è molto significativo ($p < 0,0000001$), confermando la fortissima correlazione fra le variabili osservate. La bontà della soluzione identificata è dimostrata anche dal valore assunto dall'indice KMO (KMO=0,917), che accerta un'ottima adeguatezza campionaria.

Tabella 1. Autovalori delle componenti principali estratte e dei fattori ruotati con Promax.

Componente	Soluzione iniziale			Soluzione ruotata ^a
	Autovalori	% varianza	% cumulata	Autovalori
1	9,596	38,38	38,38	7,516
2	2,354	9,42	47,80	6,693
3	1,734	6,94	54,74	7,119
4	1,521	6,09	60,82	4,887
5	1,207	4,83	65,65	2,462
6	1,142	4,57	70,22	3,355
7	0,773	3,09	73,31	
...	
25	0,081	0,32	100,00	

a Quando i fattori sono correlati a causa di una rotazione non ortogonale, gli autovalori non possono essere sommati per ottenere stime della rispettiva quota di varianza totale.

³ L'Analisi delle Componenti Principali (ACP) viene effettuata per studiare, riassumere e semplificare un insieme di variabili osservate, riducendo la complessità dei dati tramite combinazioni lineari di tali variabili sotto la condizione che ad ogni C.P., dalla prima all'ultima, venga attribuito il *massimo possibile* della variabilità del sistema, e che inoltre esse siano *indipendenti* l'una dall'altra. Il risultato è che, in genere, poche C.P. contengono (in forma modificata) la maggior parte delle informazioni contenute in insiemi decisamente più ampi di variabili osservate. Tuttavia, tale procedura è poco adeguata a variabili ordinali (richiedendo che i residui abbiano distribuzione normale), e quindi per l'estrazione delle componenti è stato qui utilizzato il metodo CATPCA (*Categorical Principal Component Analysis*), sviluppato dal *Data Theory Scaling System Group* dell'Università di Leida, NL (De Leeuw *et al.*, 1976; Meulman & Heiser, 1999; Meulman *et al.*, 2004). Tale metodo, derivato dall'Analisi delle Corrispondenze Multiple, appartiene alla famiglia *PrincAls*, basata sull'algoritmo ALSOS (*Alternative Least Squares Optimal Scaling*), e consente di utilizzare sia variabili scalari che ordinali e perfino categoriali non ordinabili.

Per avere migliori indicazioni su tali variabili, partendo dalle componenti principali identificate si è poi avviata una analisi fattoriale tramite rotazione non ortogonale *Promax*⁴, al fine di ottenere una soluzione più semplice e adeguata alla migliore interpretabilità dei risultati senza rinunciare ad investigare su eventuali relazioni tra i fattori. La rotazione ha permesso così di individuare, all'interno dei sei fattori (dimensioni latenti della qualità del sito), le variabili maggiormente caratterizzanti ciascuna dimensione.

Nella Tabella 2, oltre alle comunanze degli *item* osservati (ossia la quota della loro variabilità che viene spiegata dal sistema fattoriale, e dunque l'importanza che ciascuna variabile osservata ha entro il sistema fattoriale) sono appunto riportati i rispettivi "factor loadings", valori che esprimono l'intensità della relazione tra le variabili osservate ed i fattori ottenuti, e che risultano indispensabili per comprendere e interpretare in senso estensivo tali fattori⁵.

In base al peso rivestito dagli item considerati, i fattori evidenziano quindi le dimensioni fondamentali che sottendono il giudizio degli utenti.

Il primo fattore (*Accessibilità e Usabilità*) risulta principalmente correlato alle seguenti variabili:

- Chiarezza mappa del sito;
- Localizzazione informazioni con pochi click;
- Accessibilità mappa;
- Classificazione delle categorie durante la navigazione.

Il secondo fattore (*Velocità di accesso*) è correlato alle seguenti variabili:

- Tempo di apertura delle pagine;
- Tempo di caricamento del sito;
- Tempo di download;

⁴ L'Analisi Fattoriale (A.F.) ha lo scopo di chiarire relazioni precedentemente identificate tra un insieme di variabili osservate e un insieme di componenti o fattori, ossia dimensioni latenti in grado di render conto delle similarità che accomunano una serie di variabili (ritenendo che detti fattori riflettano strutture o processi latenti che sono all'origine delle "similarità" identificate). L'A.F. si basa, essenzialmente, su una rotazione matematica dei coefficienti di tali variabili latenti nello spazio fattoriale k-dimensionale fino al punto che meglio rappresenta i criteri di ottimizzazione prestabiliti. Esistono rotazioni ortogonali, come il metodo VARIMAX (Kaiser, 1958), e non ortogonali. Lo scopo principale delle rotazioni oblique come PROMAX (Manly, 1986) è di identificare eventuali correlazioni tra fattori, nonché di identificare meglio i fattori medesimi, incrementandone il peso per alcune variabili e riducendolo per altre, con il risultato di esaltare l'effetto di tutti gli elementi con modulo maggiore di uno e smorzare l'effetto di quelli con modulo minore di uno. Il vantaggio di questo metodo, la rotazione obliqua, sta nel fatto che, se i fattori ruotati restano ancora tendenzialmente ortogonali, si può essere sicuri che tale ortogonalità dipenda dalla rotazione, ma esclusivamente dalla struttura fattoriale. Tale criterio è, come gli altri metodi di rotazione, indipendente dal metodo di estrazione scelto.

⁵ Allo scopo di rendere più chiara la relazione tra una variabile ed i fattori, in dette tabelle sono omessi i valori inferiori a 0,33, comunque forniti dall'analisi.

- Velocità dello scrolling;
 - Visualizzazione del sito su qualsiasi browser.
- Il terzo fattore (*Informazioni e contenuti*) è correlato soprattutto a:
- Comprensibilità del lessico utilizzato;
 - Livello di utilità delle informazioni pubblicate;
 - Chiarezza dei contenuti;
 - Livello di approfondimento e dettaglio dei contenuti;

Tabella 2. *Pesi fattoriali e comunanze degli item della soluzione ruotata con Promax.*

Item	Fattori						Comunanze
	F1	F2	F3	F4	F5	F6	
Chiarezza mappa del sito	0,949						0,793
Localizzazione informazioni con pochi click	0,918						0,785
Accessibilità mappa	0,857						0,696
Classificazione delle categorie durante la navigazione	0,822						0,705
Terminologia comprensibile	0,683						0,555
Informazioni utili presenti sul sito	0,521		0,350				0,604
Servizi offerti/Semplificazione attività	0,482		0,366				0,571
Tempo di apertura delle pagine		0,910					0,839
Tempo di caricamento del sito		0,908					0,815
Tempo di download		0,879					0,776
Velocità dello scrolling		0,836					0,758
Visualizzazione sito su ogni browser		0,827					0,705
Comprensibilità del lessico utilizzato			0,868				0,699
Livello utilità informazioni pubblicate			0,849				0,703
Chiarezza dei contenuti			0,809				0,730
Livello di approfondimento e dettaglio dei contenuti			0,795				0,713
Adeguatezza del contrasto tra colori caratteri e colore sfondo				0,855			0,776
Dimensione dei caratteri				0,808			0,733
Visibilità dei tratti distintivi del sito				0,712			0,556
Scelta della lingua					0,890		0,760
Tempestività risposte/segnalazione disservizi tecnici nei Contact Form					0,789		0,633
Accuratezza/correttezza traduzione					0,648		0,606
Messaggio errore/Azione correttiva						0,891	0,811
Segnalazione errori o malfunzionamenti durante la navigazione						0,785	0,640
Errore/recupero dati						0,659	0,593

Il quarto fattore (*Grafica e leggibilità*) è correlato a:

- Adeguatezza del contrasto tra colori caratteri e colore sfondo;
- Dimensione dei caratteri;
- Visibilità dei tratti distintivi del sito.

Il quinto fattore (*Interazioni*) è correlato a:

- Scelta della lingua;
- Tempestività risposte/segnalazione disservizi tecnici con Form di contatto;
- Accuratezza e correttezza della traduzione.

Il sesto fattore (*Gestione errori*) è correlato a:

- Messaggio errore/Azione correttiva;
- Segnalazione errori o malfunzionamenti durante la navigazione;
- Errore/recupero dati.

Dalla tabella, nella quale sono stati riportati i valori che esprimono l'intensità della relazione tra le variabili ed i fattori considerati, è possibile inoltre notare come alcune variabili esercitino la loro influenza contemporaneamente in più fattori:

Le variabili *Utilità delle informazioni presenti sul sito* e *Servizi offerti/Semplificazione attività* influenzano in modo significativo sia il Fattore 1 (*Accessibilità e Usabilità*) che il Fattore 3 (*Informazioni e contenuti*);

La Tab. 3 esplicita le correlazioni residue tra i fattori, che è possibile attribuire all'influenza reciproca delle dimensioni latenti, e non alle correlazioni tra le variabili originarie (annullate dalla iniziale ACP). Risulta evidente la forte correlazione del fattore di *Accessibilità e Usabilità* con gli altri, pur se in modo meno rilevante con il fattore *Interazioni*. Il secondo fattore (*Velocità di accesso*) si correla fortemente con il fattore 3 (*Informazioni e contenuti*) e con la *Grafica e leggibilità*, e un po' meno con *Gestione errori* del sito, mentre il terzo risulta connesso ai questi due ultimi in modo più cospicuo del precedente. Appare evidente una struttura di relazioni molto articolata, con coefficienti di correlazione molto elevati fra i vari fattori, in particolare i primi tre.

Tabella 3. Correlazioni tra i fattori identificati con rotazione non ortogonale Promax.

Fattori	<i>Velocità di accesso</i>	<i>Informazioni e contenuti</i>	<i>Grafica e leggibilità</i>	<i>Interazioni</i>	<i>Gestione errori</i>
<i>Accessibilità e Usabilità</i>	0,480	0,628	0,434	0,270	0,397
<i>Velocità di accesso</i>		0,553	0,474	0,089	0,282
<i>Informazioni e contenuti</i>			0,496	0,187	0,343
<i>Grafica e leggibilità</i>				0,084	0,183
<i>Interazioni</i>					0,104

4. Mappa di prossimità delle variabili osservate

Le variabili selezionate dall'AF, essendo in effetti item qualitativi ordinali, sono state analizzate anche tramite una procedura ottimizzata di Analisi delle Corrispondenze Multiple⁶, allo scopo sia di confermare le similarità identificate dalla struttura fattoriale e sia di identificare percorsi di relazioni privilegiate, tramite mappa visuale. Tale procedura, definendo le prossimità fra i baricentri delle variabili osservate, è utile anche per chiarire le relazioni tra i fattori a cui tali variabili sono connesse.

L'analisi delle corrispondenze multiple quantifica i dati categoriali, anche nominali, assegnando valori numerici ai casi e alle categorie, in modo che i casi all'interno della stessa categoria siano vicini tra loro e gli oggetti in diverse categorie siano distanti. Ciascun caso si trova dunque il più vicino possibile ai punti delle categorie a esso applicabili, e in tal modo i casi sono suddivisi in sottogruppi omogenei rispetto alle categorie utilizzate.

Essa parte da una "matrice-indicatore", o "disgiuntiva completa", che è una matrice composta solo dalle cifre 0 e 1: per ogni variabile x_j si costruisce innanzitutto uno scalare g_{jh} che assume il valore 1 oppure 0 a seconda che l' i -esimo individuo rientri o non rientri nella h -esima categoria della variabile; il vettore g_{jh} è dato da tale scalare esteso a tutti gli individui in relazione alla categoria h di x_j . Considerando tutte le categorie di x_j , i vettori colonna g_{jh} originano la matrice-indicatore G_j di dimensioni $n \times k_j$. Estendendo la procedura a tutte le m variabili categoriali otteniamo la matrice *disgiuntiva completa* $G = [G_1 \dots G_j \dots G_m]$, di ordine $n \times K$ (ove $K = \sum_j k_j$). In tal modo, ogni variabile categoriale osservata viene specificata come prodotto di una matrice-indicatore (che è nota, essendo fornita dai valori osservati) e di un vettore $\beta_j = [\beta_{j1} \dots \beta_{jh} \dots \beta_{jk_j}]'$ di parametri di scaling.

La procedura di ACM qui utilizzata appartiene, come la tecnica CATPCA utilizzata nel paragrafo precedente, alla classe delle procedure di analisi ottimale sviluppate dal *Data Theory Scaling System Group* della Leiden University, basate sul metodo dei Minimi Quadrati Alternati (Gifi, 1980). In particolare, l'algoritmo di Optimal ACM (De Leeuw & Van Rijkevorsel, 1980) effettua per prima cosa, in modo casuale, una stima iniziale X^* della matrice di variabili X che soddisfi i

⁶ L'Analisi delle Corrispondenze Multiple, nata come estensione dell'Analisi delle Corrispondenze inizialmente proposta da Hirschfeld (1935) e poi sviluppata da Benzécri e dai suoi collaboratori nel 1973, rappresenta uno strumento molto utile per lo studio delle relazioni tra p caratteri statistici qualitativi, ognuno caratterizzato da m_j modalità ($j=1, \dots, p$), ed è in grado di identificare strutture latenti in un insieme di dati, rappresentandoli come punti-baricentro in uno spazio euclideo k dimensionale ($k < p$). Cfr. Greenacre, 1988, 2006.

vincoli di calcolo (che non vi sia alcuna influenza di dati fuori dall'intervallo ammesso e che i punteggi assegnati siano centrati), e stima la relativa funzione di perdita tramite appunto la matrice disgiuntiva completa \mathbf{G} . Poi, tramite vari passaggi (basati anche su una prima quantificazione "casuale" delle categorie) si stima nuovamente la \mathbf{X}^* in modo che rispetti, oltre ai predetti vincoli, la condizione aggiuntiva della maggior prossimità (con metodo dei minimi quadrati), a una matrice empirica ricavata anch'essa tramite la matrice \mathbf{G} . Aggiornando la quantificazione delle categorie in base a questa nuova stima, si calcola una nuova funzione di perdita, e poi, se non si è raggiunta la convergenza a un criterio predefinito, si stima una nuova \mathbf{X}^* , e il procedimento prosegue fino a convergenza⁷.

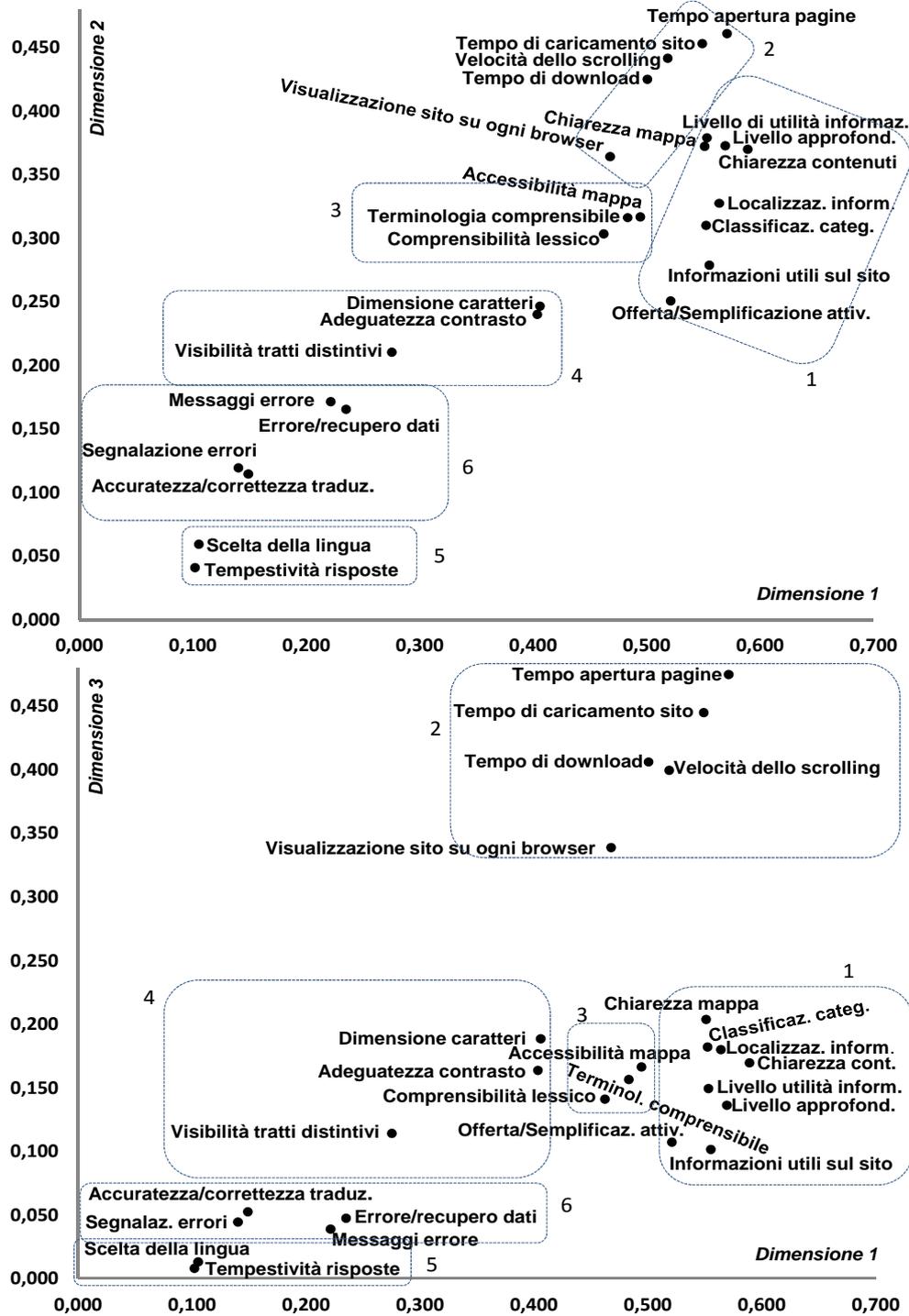
Nella presente analisi sono state definite tre dimensioni principali, che spiegano oltre l'88% dell'inerzia, ossia della variabilità della matrice totale, ma va detto che alle prime due dimensioni (che presentano un indice di Cronbach pari o superiore a 0,9 e dunque un'ottima affidabilità, mentre la terza ha un indice ancora elevato, ma uguale a solo 0,806) compete ben il 70,6% dell'inerzia totale. Per questo motivo, la mappa grafica della Figura 1 risulta ben più chiara nell'incrocio fra le prime due dimensioni che non fra la prima e la terza (la mappa tra seconda e terza dimensione è residuale ed è quindi stata omessa).

La considerazione più interessante che si trae dalla Figura 1 è la somiglianza della mappa delle distanze fra le variabili con la loro struttura fattoriale (ruotata) riportata in Tabella 2. Infatti, nella mappa relativa alle prime due dimensioni (ove peraltro gli item sono distribuiti grossomodo attorno alla bisettrice del piano cartesiano), è possibile identificare alcuni gruppi di variabili molto prossime fra loro, e in particolare (partendo dall'origine degli assi), una coppia che corrisponde agli item distintivi del 5° fattore (*Interazioni*), poi la terna che distingue il 6° fattore (*Gestione errori*), più il terzo item del 5° fattore, e quella che caratterizza il 4° fattore (*Grafica e leggibilità*). Segue poi un ammasso di item (vagamente separate in diagonale), che competono sia al 1° (*Accessibilità e Usabilità*) che al 3° fattore (*Informazioni e contenuti*), e infine, nel punto più distante dall'origine degli assi, le 5 variabili che contraddistinguono in Fattore 2 (*Velocità di accesso*).

Poiché anche nella mappa fra prima e terza dimensione gli item caratterizzanti primo e terzo fattore appaiono molto vicini gli uni dagli altri, si può affermare che la correlazione sussistente fra detti fattori (cfr. Tabella 3) si riflette in modo chiaro anche sulle variabili che vi sono più connesse, le quali dunque descrivono giudizi statisticamente molto simili fra loro da parte di ciascun utente.

⁷ Per una descrizione più esauriente del procedimento qui sintetizzato, cfr. d'Ovidio 2012.

Figura 1. *Mappa delle corrispondenze multiple dei giudizi (prime tre dimensioni).*



Inoltre, la distanza crescente di tali gruppi dall'origine degli assi sembra suggerire una interessante gerarchia di importanza dei fattori latenti che a questi possono assimilarsi, e dunque la maggiore importanza viene assunta dal fattore “*Velocità di accesso*”, seguita da “*Accessibilità e Usabilità*” e “*Informazioni e contenuti*” e dal fattore “*Grafica e leggibilità*”; a distanza, con livelli di importanza meno rilevanti, troviamo “*Gestione errori*” e “*Interazioni*”, presumibilmente perché gli eventi che danno luogo ad accessi per questi motivi sono meno frequenti (e dunque interessano meno utenti) rispetto alla ricerca di informazioni che rientrano negli altri gruppi.

In ogni caso, dando per acquisita la notevole differenza concettuale e matematica tra ACM e AF (soprattutto tenendo conto della rotazione effettuata in questa), il netto parallelismo dei risultati conferma la robustezza dei medesimi e consente di ipotizzare una stabile struttura sottostante.

5. Dimensioni della qualità ed elementi di fidelizzazione

Come ultimo step della presente disamina di dati, si è ritenuto necessario innanzitutto classificare, tramite analisi di segmentazione, gli utenti in base sia ai fattori identificati e sia alle caratteristiche risultate significative nel corso dell'analisi esplorativa in riferimento alla proxy “Accesso al sito web” (Tipo di corso di studi, Pseudo-facoltà di afferenza del corso, Anni di iscrizione; i vari motivi di connessione al sito web sono stati invece esclusi a causa della relazione fin troppo evidente e “banale” che sussiste tra essi e la frequenza di accesso).

L'analisi di segmentazione (con algoritmi: CHAID e C@RT), la quale ha lo scopo appunto di raggruppare fra loro le unità statistiche in gruppi (nodi) il più possibile omogenei al loro interno in base alla relazione con una variabile di interesse (variabile oggetto), è stata effettuata su un campione casuale pari al 50% del collettivo osservato (“campione di training”), applicando poi le regole di decisione ottenute a un “campione di test” comprendente il restante 50% (cfr., ad es, Breiman *et al*, 1984). Per massimizzare la capacità discriminante della procedura, la variabile “Accesso al sito web” è stata resa binaria, raggruppando nella categoria “Poco o periodicamente” gli accessi con frequenza occasionale o non oltre una volta a settimana (59% dei casi), e nella categoria “Molto frequentemente” quelli con frequenza più volte alla settimana o più volte al giorno (41%).

Sono stati utilizzati gli usuali criteri di arresto: massimo numero di livelli pari a 5 per CHAID e a 8 per C@RT (con procedura di *pruning*), minima numerosità dei nodi figli (ossia quelli che si ottengono da ogni suddivisione) pari a 30 e minima numerosità dei nodi genitori (quelli che vengono suddivisi) pari a 60.

I risultati migliori e più stabili nel confronto fra campione di training e campione di test si ottengono applicando l'algoritmo binario C@RT, e, pur non essendo ottimi, appaiono singolarmente molto stabili e parsimoniosi: invero, in due soli livelli di segmentazione si ottiene una capacità discriminante con rischio di errata classificazione pari a 0,37 nel campione di training e pari a 0,38 nel campione di test, con $\text{std.err.}=0,02$ in ambo i casi (Tabella 4).

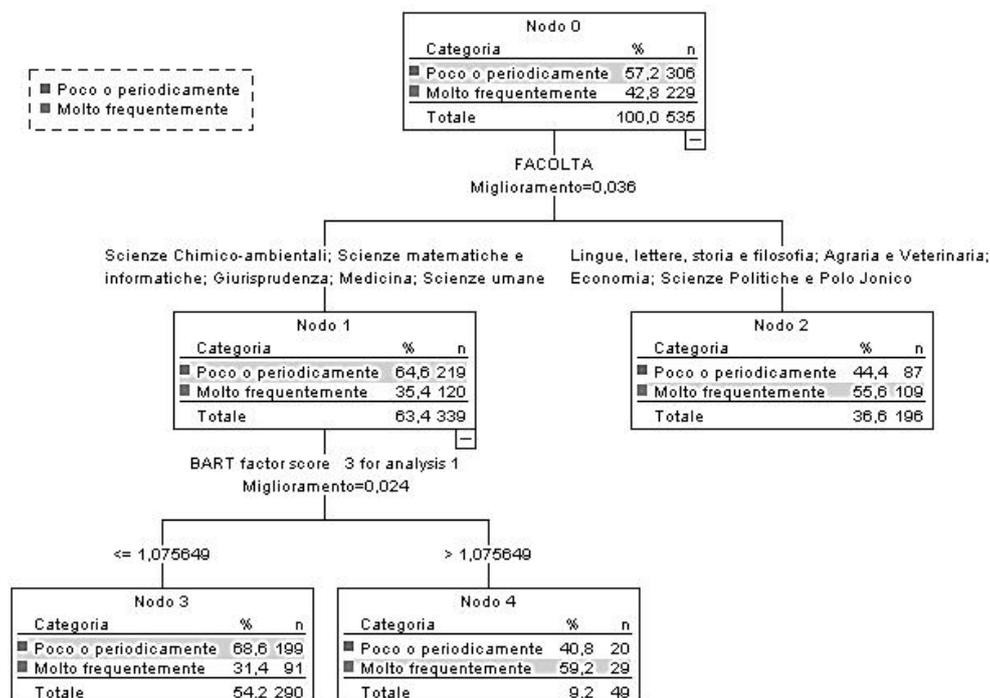
Tabella 4. Matrice di confusione relativa all'analisi di segmentazione

<i>Frequenza di accesso al sito OSSERVATA</i>		<i>Frequenza di accesso al sito PREVISTA</i>		
		<i>Poco o periodicamente</i>	<i>Molto frequentemente</i>	<i>% di corretta classificazione</i>
Training	<i>Poco o periodicamente</i>	199	107	65,0
	<i>Molto frequentemente</i>	91	138	60,3
	Totale %	54,2	45,8	63,0
Test	<i>Poco o periodicamente</i>	186	128	59,2
	<i>Molto frequentemente</i>	69	131	65,5
	Totale %	49,6	50,4	61,7

Le regole di classificazione identificate riescono a prevedere in modo abbastanza simile nei campioni di training e di test, ma per via della casualità di estrazione dei campioni con modalità leggermente diverse: nel primo, invero, vengono previsti meglio gli accessi meno frequenti (65% dei quali vengono correttamente classificati), mentre nel campione di test sono gli accessi più frequenti ad essere meglio previsti (65,5% di casi correttamente classificati). L'albero di classificazione ottenuto è riportato in Figura 2, e come già affermato è molto parsimonioso, esplicando la propria azione in soli 2 livelli (escludendo il nodo di partenza) e in 3 nodi finali, cioè non più suddivisibili in base ai criteri prefissati).

Il primo e più importante criterio discriminante è la "Facoltà" di afferenza del corso di studi, posto che nel gruppo degli studenti di Lingue, lettere, storia e filosofia, di Agraria e Veterinaria, di Economia, di Scienze Politiche e del Polo Jonico si registra la maggior percentuale di accessi molto frequenti. Tale nodo non si divide oltre (o, per meglio dire, le sue ulteriori suddivisioni non risultano abbastanza importanti nell'economia generale, per cui la procedura di *pruning* le elimina automaticamente), mentre quello composto dagli studenti delle altre Facoltà risulta soggetto al 3° Fattore (*Informazioni e contenuti*), che ove superiore alla soglia di 1,076 porta a una maggior quota di accessi molto frequenti, mentre se inferiore a tale soglia si rilevano soprattutto accessi meno frequenti. Si tenga conto, peraltro, che il fattore è distribuito all'incirca come una v.c. normale standardizzata.

Figura 2. Albero di classificazione relativo all'accesso frequente al sito web uniba.it.



6. Considerazioni finali

L'analisi della qualità di un sito web non può certamente limitarsi ai pochi aspetti descritti nelle pagine precedenti, ma il presente studio va considerato un primo approccio al problema, con un occhio particolare alle particolarità dell'Università degli Studi di Bari Aldo Moro, università generalista fra le più grandi e articolate del Paese. Ulteriori sviluppi sono infatti previsti (ad es., analisi strutturale/causale).

La cifra del presente studio è data soprattutto dalle differenti gerarchie di importanza delle variabili studiate, in particolare le dimensioni latenti ottenute dall'analisi fattoriale, delle quali la più importante per il concetto di qualità del sito sembra essere l'*Accessibilità e usabilità*, seguita dalla *Velocità di accesso* (entrambe dimensioni più "tecniche", e dunque affidate al design del sito e all'efficienza delle routine di navigazione) e solo in terza battuta dalla vera ragion d'essere di un sito web universitario, fornire *Informazioni e contenuti*. Tale osservazione risulta, però, molto ridimensionata se posta in relazione alla proxy di fidelizzazione, nel cui discrimine entrano solo la Facoltà di afferenza del corso di studi e, appunto, il fattore *Informazioni e contenuti*, a cui viene dunque resa giustizia.

Bibliografia

- Benzécri, J.P. et al. (1973). *L'Analyse des données, Tome II: L'Analyse des correspondances*. Dunod éd., Paris.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J., 1984. *Classification and Regression Trees*. Chapman & Hall, New York-London.
- Browne, M. W. (1982). Covariance structures. In: D.M. Hawkins (ed.), *Topics in applied multivariate analysis*, Cambridge University Press: 72-141.
- De Leeuw J., Young F.W., Takane Y., 1976. Additive Structure in Qualitative Data: an Alternative Least Squares Method with Optimal Scaling Features. *Psychometrika* 41, 471-504.
- De Leeuw, J., Van Rijkevorsel, J. (1980). Homals and Princals - Some Generalizations of Components Analysis. In: E. Diday, Y. Escoufier, L. Lebart, J.P. Pages, Y. Schektman, R. Tomassone (eds.), *Data Analysis and Informatics*, Amsterdam (NL): 231-241.
- Delvecchio, F. (2015). *Statistica per l'analisi dei fenomeni sociali*, Cleup, Padova.
- d'Ovidio, F (ed.) (2012). *Elementi di Statistica per la valutazione dei servizi*, Cleup, Padova.
- Gifi, A. (1981). *Nonlinear Multivariate Analysis*, Department of Data Theory, University of Leiden, NL.
- Greenacre, M. J. (1988). Correspondence analysis of multivariate categorical data by weighted least squares. *Biometrika*, **75**: 457-467.
- Greenacre, M. J. (2006). From simple to multiple correspondence analysis. In M. J. Greenacre, J. Blasius (eds.), *Multiple correspondence analysis and related methods*. Dordrecht: Chapman and Hall (Kluwer): 41-76.
- Hirschfeld, H.O. (1935). A connection between correlation and contingency, *Proc. Cambridge Philosophical Society*, **31**: 520-524
- Jöreskog, K. G. (1977). Structural equation models in the social sciences. In: Krishnaiyah, P.R. (ed.) *Application of Statistics*. North Holland, Amsterdam: 265-287.
- Kaiser H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, **23**: 187-200.
- Manly, B.F.J. (1986) *Multivariate Statistical Methods: a Primer*, Chapman & Hall, London.
- Meulman J.J., Heiser W.J., 1999. *Categories 10.0*. SPSS Inc., Chicago.
- Meulman J. J., Van der Kooij A. J., Heiser W. J., 2004. Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In: Kaplan D. (ed.), *Handbook of quantitative methodology for the social sciences*, Sage, London, 49-70.
- Yoo, S., Jin, J. (2004). Evaluation of the home page of the top 100 university web sites, *Academy of Information and Management Sciences*, 8 (2): 57-69.
- Zhang, P., Dran, G. (2001). Expectations and ranking of website quality features: results of two studies on user perceptions. In: *Proceedings of the 34th Hawaii International Conference on System Sciences*.