



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede amministrativa: Università degli Studi di Padova

Dipartimento di Psicologia Generale

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE PSICOLOGICHE
INDIRIZZO PSICOLOGIA SPERIMENTALE
CICLO XXIV

Criteri dell'Informazione e Selezione dei Modelli in Misurazione Funzionale

Direttore della Scuola: *Ch.ma prof.ssa Clara Casco*

Coordinatore d'Indirizzo: *prof.ssa Francesca Pazzaglia*

Supervisore: *Ch.mo prof. Giulio Vidotto*

Dottorando: *Davide Massidda*

Indice

Riassunto	1
Abstract	9
1 La Teoria dell'Integrazione delle Informazioni	17
1.1 Introduzione	17
1.2 Il processo di formulazione del giudizio	18
1.3 L'Algebra Cognitiva	21
1.3.1 I modelli additivo e moltiplicativo	22
1.3.2 Il modello averaging	23
1.4 Il concetto di peso	26
1.5 La Misurazione Funzionale	28
1.6 L'identificazione del modello	29
1.6.1 Modelli additivi e moltiplicativi	29
1.6.2 Modelli averaging	31
1.7 Il modello averaging a pesi diversi	33
1.7.1 Identificabilità e unicità	35
2 Il metodo R-Average	37
2.1 Introduzione	37
2.2 La selezione dei modelli basata sui criteri dell'informazione	39
2.3 Struttura del metodo	41
2.3.1 Routine di minimizzazione	41
2.3.2 Procedure per la stima e la selezione del modello	41
2.3.3 L'identificazione delle uguaglianze fra i pesi	44

2.4	Applicazione del metodo	45
2.5	Differenze fra R-Average e AVERAGE	46
2.6	Verifica dell'attendibilità: studio Monte Carlo	47
2.6.1	Metodo	48
2.6.2	Risultati	51
2.6.3	Discussione	52
2.7	La riformulazione del modello	53
2.8	Nuove procedure di stima	56
2.9	Nuovo studio Monte Carlo	58
2.9.1	Metodo	58
2.9.2	Risultati	58
2.9.3	Discussione	61
3	Il modello averaging nelle ricerche di marketing	63
3.1	Introduzione	63
3.2	Metodo	65
3.2.1	Partecipanti	66
3.2.2	Apparato e stimoli	66
3.2.3	Procedura	67
3.3	Risultati	68
3.4	Discussione	70
4	Verifica dell'ipotesi averaging nell'effetto di subadditività	73
4.1	Introduzione	73
4.2	Possibili spiegazioni del fenomeno	74
4.3	Lo studio proposto	77
4.4	Metodo	78
4.4.1	Partecipanti	78
4.4.2	Disegno sperimentale	78
4.4.3	Ipotesi	79
4.4.4	Apparato e stimoli	80
4.4.5	Procedura	81
4.5	Risultati	82

4.5.1	Manipolazione della fiducia	82
4.5.2	Validità del modello averaging	84
4.5.3	Linearità della funzione di risposta	88
4.6	Discussione	89
5	Il modello averaging nella valutazione delle credenze di fiducia	91
5.1	Introduzione	91
5.2	Metodo	94
5.2.1	Partecipanti	94
5.2.2	Disegno sperimentale	94
5.2.3	Apparato e stimoli	94
5.2.4	Procedura	95
5.3	Risultati	96
5.4	Discussione	98
	Conclusion	101
	A Risultati degli studi Monte Carlo	103
	B Implementazione del metodo R-Average	119
B.1	La funzione <i>rav</i> per la stima dei modelli averaging	120
B.2	Principali funzioni di supporto all'analisi	122
	C Credenze di fiducia: item proposti	123
C.1	Ambito individuale	123
C.2	Ambito istituzionale	124
C.3	Ambito organizzativo	125
	Bibliografia	127

Riassunto

I processi di valutazione degli stimoli ambientali e di decisione sono comuni nella vita quotidiana e in tante situazioni di carattere sociale ed economico. Questi processi sono generalmente descritti dalla letteratura scientifica utilizzando modelli di scelta multi-attributo. Tali modelli assumono che la valutazione di uno stimolo descritto da più attributi sia il risultato di un processo a più stadi (Anderson, 1981; Lynch, 1985): valutazione degli attributi, integrazione dei valori e valutazione esplicita dello stimolo. Comunemente, in questo campo, le situazioni sperimentali richiedono la valutazione di un set di stimoli costruiti combinando diversi attributi. Un soggetto valutatore esamina gli attributi di ogni stimolo; usando il solo modello “mentale” di scelta, assegna un valore agli attributi e formula un giudizio globale. Infine, il soggetto esprime la sua opinione in termini di ordinamento, preferenze a coppie, valori su una scala numerica e così via.

Questa cosiddetta valutazione multi-attributo soffre di una fondamentale difficoltà nel misurare i valori di ogni attributo di uno stimolo partendo dalle valutazioni complessive di ogni soggetto. Fondamentalmente, il problema è derivare ogni valore decomponendo il giudizio complessivo (cioè la risposta in output). Questa difficoltà di misurazione è tipica di molte delle spesso complementari tradizioni dei modelli multi-attributo, come la Conjoint Analysis (Luce e Tukey, 1964; Green e Rao, 1971; Krantz e Tversky, 1971) o la Teoria dell’Integrazione delle Informazioni (IIT: Anderson, 1970, 1981, 1982).

Secondo la IIT di Anderson, il sistema cognitivo fornisce un valore soggettivo a ogni caratteristica di uno stimolo, e tali valori vengono combinati in un giudizio complessivo utilizzando una specifica funzione d’integrazione. La IIT descrive le modalità d’integrazione utilizzando differenti regole matema-

tiche, e la misurazione funzionale è la metodologia proposta per determinare e misurare la funzione d'integrazione. La misurazione funzionale si serve di esperimenti fattoriali, selezionando alcuni attributi di uno stimolo e combinandoli in piani fattoriali. Solitamente, le valutazioni dei soggetti per ogni cella del disegno sperimentale sono riportate su una category scale, e ogni soggetto ripete la valutazione per più prove. Partendo dalle valutazioni soggettive, la misurazione funzionale mira a quantificare il valore di ogni livello dei fattori e la sua importanza nel giudizio complessivo, per ogni soggetto valutatore o gruppo di soggetti.

La teoria di Anderson suggerisce che le regole d'integrazione più ampiamente utilizzate sono di tre fondamentali e semplici tipologie: additiva, moltiplicativa e di media ponderata (averaging). Tecniche statistiche come l'analisi della varianza possono essere utilizzate per individuare la regola d'integrazione sulla base della bontà dell'adattamento.

La regola averaging in particolare è in grado di tenere in considerazione gli effetti d'interazione tra i fattori, scindendo la valutazione in due componenti: valore di scala e peso, che possono essere identificati e misurati separatamente (Zalinski e Anderson, 1989). Se il valore di scala rappresenta il posizionamento del livello del fattore sulla scala di risposta, il peso rappresenta la sua importanza nel giudizio complessivo. Il modello averaging fornisce una via molto utile per gestire gli effetti d'interazione tra i fattori, superando l'assunto d'indipendenza sul quale molte applicazioni dei modelli di scelta multi-attributo sono basate. Tuttavia, il modello presenta alcuni punti critici relativi alla questione della stima, e per questo motivo il suo potenziale non è stato pienamente sfruttato fin'ora. In questo lavoro di ricerca viene proposto un nuovo metodo per la stima dei parametri del modello averaging. Il metodo consente di selezionare il miglior set di parametri per adattare i dati, e mira a superare alcuni problemi che ne hanno limitato l'uso.

Secondo questo nuovo metodo, chiamato R-Average (Vidotto e Vicentini, 2007; Vidotto, Massidda e Noventa, 2010), la scelta del miglior modello è fatta in accordo al cosiddetto "principio di parsimonia": il miglior modello è quello più "semplice", che trova il miglior compromesso tra spiegazione del fenomeno (varianza spiegata) e complessità strutturale (numero di parametri

di peso diversi). Il processo di selezione usa in combinazione due indici di bontà dell'adattamento: l'Akaike Information Criterion (AIC; Akaike, 1974) e il Bayesian Information Criterion (BIC; Schwartz, 1978). Entrambi gli indici sono ricavati partendo dal logaritmo della varianza residua pesata per il numero di osservazioni, e penalizzando i modelli con parametri aggiuntivi. AIC e BIC differiscono nella funzione di penalizzazione - dato che il BIC impone una penalità maggiore ai modelli con più parametri - e sono molto utili per la comparazione fra modelli.

In questo lavoro di ricerca vengono presentate due versioni del metodo R-Average. Queste due versioni sono una l'evoluzione dell'altra, ed entrambi i metodi sono strutturati in diverse procedure per eseguire la stima. Fondamentalmente, R-Average consta di tre procedure: procedura EAM, procedura DAM e procedura Information Criteria (IC). EAM, DAM e IC differiscono nei vincoli imposti sui pesi durante il processo di ottimizzazione.

La procedura EAM vincola tutti i pesi all'interno di ogni fattore a essere uguali, stimando un modello a pesi uguali. Questo modello è il migliore in termini di parsimonia, perché presenta il minor numero di parametri (uno unico per ogni fattore). Infatti, si definisce come "parsimonioso" un modello semplice, nel quale i pesi sono uguali. Diversamente, la procedura DAM non impone alcun vincolo sui pesi, lasciandoli liberi di variare. Così, questa procedura può potenzialmente convergere verso un modello averaging a pesi completamente diversi (dove cioè tutti i pesi dei livelli di ogni fattore sono diversi).

Il cuore del metodo R-Average è la procedura Information Criteria. Questa procedura è basata sull'idea che, da un punto di vista psicologico, un modello semplice è più plausibile di un modello complesso. Per questo motivo, l'algoritmo di stima non è volto alla ricerca dei parametri che spiegano la maggior quota di varianza, ma cerca un compromesso tra varianza spiegata e complessità del modello. Un modello complesso sarà valutato come migliore di uno più semplice solo se permette di ottenere un grado significativamente superiore di spiegazione del fenomeno.

La procedura IC cerca il modello provando a tenere (nella versione "forward") o a rendere (nella versione "backward") tutti i pesi uguali. Nella

versione forward, la procedura parte dal modello EAM e passa in rassegna tutte le possibili combinazioni di pesi, modificandole: inizialmente uno a uno, poi due a due, poi tre a tre e così via. Per ogni combinazione, la procedura prova a diversificare i pesi. Di volta in volta, utilizzando gli indici BIC e AIC, la procedura seleziona il miglior set di parametri e assume il modello selezionato come riferimento per il passo successivo (se un'evidenza di miglioramento viene trovata). Nella versione backward, la procedura parte dal modello DAM e passa in rassegna tutte le possibili combinazioni di pesi, provando a renderli uguali. Gli indici BIC e AIC sono utilizzati per comparare i nuovi modelli con quelli di riferimento: se un nuovo modello viene individuato come migliore di quello di riferimento, sarà utilizzato come nuovo riferimento per i passi successivi.

Infine, tutti i modelli stimati dalle procedure vengono comparati, e il quello migliore sulla base dei criteri dell'informazione viene scelto.

La formulazione originale del modello averaging è stata modificata nell'evoluzione del metodo R-Average di base. Questa riformulazione considera il peso non come semplice parametro w ma come $w = \exp(t)$. Questa trasformazione esponenziale conduce a una soluzione del classico problema di unicità che affligge la formulazione averaging (Vidotto, 2011). Inoltre, essa giustifica l'applicazione di algoritmi di cluster analysis sui parametri di peso, necessari per le procedure di raggruppamento dei soggetti sperimentali sulla base delle loro similarità. Infatti, la distanza tra due valori t può essere valutata in termini di semplice differenza. Diversamente, la distanza tra due valori w può essere valutata solo in termini di rapporto tra loro. Ciò consente l'uso di algoritmi di raggruppamento dei soggetti basati su matrici di prossimità fra i parametri.

La performance di R-Average è stata testata utilizzando studi Monte Carlo e applicazioni pratiche in tre differenti campi di ricerca: nel marketing, nella teoria delle decisioni economiche e nella fiducia interpersonale.

I risultati degli studi Monte Carlo mostrano una buona capacità del metodo di identificare i parametri del modello averaging. I parametri di scala sono in generale ben stimati. Diversamente, la stima dei pesi è un po' più critica. La stima puntuale del valore reale del peso non è precisa come quella dei valori

di scala, in particolare all'aumento della deviazione standard della componente d'errore dei dati. Nonostante questo, le stime appaiono attendibili, e le uguaglianze fra i pesi sono identificate. L'aumento del numero di replicazioni sperimentali può aiutare la selezione del modello quando gli errori presentano una grande deviazione standard. In sintesi, R-Average si configura come uno strumento molto utile per selezionare il miglior modello all'interno della famiglia dei modelli averaging, permettendo di gestire particolari condizioni multi-attributo negli esperimenti di misurazione funzionale.

Il metodo R-Average è stato applicato in un primo studio nel campo del marketing. Nell'acquistare un prodotto, le persone esprimono una preferenza per particolari prodotti: comprendere i processi cognitivi sottostanti la formulazione delle preferenze dei consumatori risulta quindi un punto importante. Lo studio è stato condotto in accordo con un produttore locale di pasta, l'azienda Sgambaro. Gli scopi della ricerca erano tre: comprendere la formulazione dei giudizi dei consumatori su un prodotto di mercato, testare il metodo R-Average in condizioni reali e fornire all'azienda Sgambaro utili informazioni per un'ottimale commercializzazione del prodotto. Sono stati manipolati due fattori: la confezione della pasta Sgambaro (scatola con finestra, scatola senza finestra e busta di plastica) e il prezzo (0.89 euro, 0.99 euro, 1.09 euro).

Le analisi sono partite considerando le valutazioni del prodotto espresse dai partecipanti: per ogni soggetto sono stati stimati i parametri del modello averaging. Dato che la popolazione dei consumatori presumibilmente non è omogenea in quanto a preferenze, il campione complessivo è stato diviso in tre gruppi (chiamati semplicemente Cluster A, Cluster B e Cluster C) attraverso un algoritmo di cluster analysis. Per entrambi i fattori Prezzo e Confezione, i diversi raggruppamenti mostrano punteggi differenti. Il Cluster A esprime giudizi che si posizionano nel centro scala, indicando come questi partecipanti non fossero particolarmente attratti dai prodotti. All'opposto, il Cluster B esprime giudizi positivi, e il Cluster C esprime giudizi generalmente negativi con l'eccezione della confezione "scatola con finestra". Per quanto concerne la confezione, si osserva che la scatola con finestra, sebbene non sia quella preferita in tutti e tre i gruppi, ha sempre valutazione positive, mentre

i giudizi per le altre confezioni variano tra i gruppi. Inoltre, se il target di potenziali consumatori per il prodotto è la popolazione generale, la scatola con finestra può essere considerata la confezione più apprezzata.

Inoltre, nel Cluster C l'ANOVA mostra un'interazione significativa tra Prezzo e Confezione. Difatti, i parametri stimati per il modello averaging mostrano che il Cluster C è generalmente influenzato da un prezzo elevato. In questo gruppo il prezzo più alto ha un peso doppio rispetto agli altri nel punteggio finale, e ciò potrebbe invalidare l'influenza positiva della confezione "scatola con finestra". È importante notare che il gruppo che è più sensibile a un prezzo alto è anche quello che presenta i punteggi di preferenza più bassi rispetto agli altri gruppi.

In un secondo esperimento, il metodo R-Average è stato applicato in uno studio nel campo delle decisioni economiche in condizioni di rischio. L'assunzione che ha mosso lo studio è che, quando una persona deve valutare una scommessa a carattere economico in una situazione rischiosa, la persona integra cognitivamente il valore economico della scommessa con quello della probabilità di vittoria. In passato, Shanteau (1974) ha mostrato che l'integrazione tra valore e probabilità è realizzata attraverso una regola moltiplicativa. Lo studio, come quello di Lynch (1979), ha sottolineato che quando la situazione concerne due scommesse simultanee, ognuna composta da un valore e una probabilità, i giudizi per la scommessa doppia sono diversi dalla somma dei giudizi espressi per le scommesse singole. Questa osservazione, denominata effetto di subadditività, viola le assunzioni della Teoria dell'Utilità Attesa.

Lo studio proposto analizza la convenienza/soddisfazione associata alle scommesse singole e doppie. Lo studio ha proposto ai partecipanti due tipologie di scommessa. Un primo gruppo di scommesse riguardava un bene (telefono cellulare) e l'altro un servizio (messaggi SMS gratis per giorno); a ogni bene/servizio era associata la probabilità di ottenerlo. Sono state definite due condizioni sperimentali. Nella prima condizione, i soggetti giudicano le scommesse considerando che i telefoni cellulari sono prodotti da una buona compagnia, e il servizio SMS è fornito da un provider inaffidabile. Nella condizione inversa, i soggetti giudicano le scommesse considerando che i telefoni

cellulari sono prodotti con bassa qualità da una compagnia inaffidabile, e il servizio SMS è fornito da un provider robusto e affidabile.

Per le scommesse doppie, è stata ipotizzata la presenza di un modello d'integrazione averaging, e i parametri del modello sono stati stimati utilizzando R-Average per ogni soggetto. I risultati mostrano che, in presenza di una scommessa doppia, l'integrazione è pienamente compatibile con un modello averaging: la corretta regola d'integrazione sembra essere quella a media ponderata e non quella additiva.

Nell'ultimo esperimento, il modello averaging e la metodologia R-Average sono state applicate a uno studio sulle credenze di fiducia in tre contesti di vita quotidiana: interpersonale, istituzionale e organizzativo. Le credenze di fiducia sono attributi positivi che si ritiene una persona debba possedere affinché ci si possa fidare di lei. Le credenze di fiducia sono fattori rilevanti perché un individuo ne consideri un altro affidabile. Esse definiscono fino a che punto chi ripone fiducia si sente sicuro nel credere che la persona su cui ripone fiducia sia affidabile. Secondo McKnight, Cummings e Chervany (1998), le credenze di fiducia più citate sono: benevolenza, competenza, onestà e prevedibilità.

L'idea sottostante lo studio proposto è che le credenze potrebbero essere integrate cognitivamente nel concetto di affidabilità attraverso un processo di ponderazione. Il metodo R-Average è stato utilizzato per identificare i parametri del modello averaging per ogni partecipante. Come principale risultato, l'analisi mostra che, in accordo con McKnight et al. (1998), le quattro credenze principali giocano un ruolo fondamentale nel giudicare la fiducia. Inoltre, in accordo con la teoria dell'integrazione delle informazioni, un modello averaging sembra spiegare le risposte individuali. La grande maggioranza dei partecipanti potrebbe essere inquadrata come caso a pesi diversi. Mentre i valori di scala mostrano un netto andamento lineare con slopes più elevati per onestà e competenza, i pesi mostrano differenze con valori medi più elevati anche per onestà e competenza. Questi risultati sono coerenti con l'idea che attributi diversi giochino un ruolo diverso nel giudizio finale: infatti, onestà e competenza sembrano rivestire un ruolo preminente, mentre la prevedibilità sembra meno rilevante. Un'altra interessante conclu-

sione riguarda l'elevato peso assunto da un basso livello di onestà; ciò sembra mostrare come una credenza connessa alla bassa onestà giochi il principale ruolo all'interno di un giudizio finale negativo. Infine, la differente inclinazione dell'andamento dei livelli degli attributi nei tre contesti situazionali suggerisce un ruolo preminente dell'onestà nelle situazioni interpersonali e della competenza nelle situazioni istituzionali.

In conclusione, la teoria dell'integrazione delle informazioni e la misurazione funzionale sembrano rappresentare un approccio interessante per comprendere la formulazione del giudizio umano. Questo lavoro di ricerca propone un nuovo metodo per stimare i parametri dei modelli averaging. Il metodo mostra una buona capacità di identificare i parametri e apre nuovi scenari nella teoria dell'integrazione delle informazioni, fornendo un buon strumento per comprendere più nel dettaglio l'integrazione averaging degli attributi.

Abstract

The processes of evaluation of environmental stimuli and decision are common in everyday life and in many social and economic situations. These processes are generally described in scientific literature using multi-attribute choice models. These models assume that evaluation of a stimulus described by several attributes results from a multi-stage process (Anderson, 1981; Lynch, 1985): evaluation of the attributes, integration of the values of the attributes and explicit evaluation of the stimulus. Commonly, in this field, experimental settings require the evaluation of a set of stimuli built combining some attributes. A subject evaluator examines the attributes of each stimulus; using her “mental” model of choice, it assigns a value to attributes and formulate an overall judgment. Finally, subject expresses his opinion in terms of order-ranking, pairwise preference comparisons, values in a rating scale, and so on.

This so-called multi-attribute evaluation suffers of a fundamental difficulty to measure the values of each attribute of a stimulus starting by the overall evaluation of each subject. Basically, the problem is to derive each value decomposing the overall judgment (i.e. the response output). This difficulty in measuring is typical in most of the often complementary multi-attribute models traditions, as those of Conjoint Analysis (Luce and Tukey, 1964; Green and Rao, 1971; Krantz and Tversky, 1971) or Information Integration Theory (IIT: Anderson, 1970, 1981, 1982).

According to Anderson’s IIT, cognitive system give a subjective value to each characteristic of a stimulus, and the values are put together in a overall judgment using a specific integration function. IIT describe integration modalities using different mathematical rules, and functional measurement is the

methodology proposed to determine and measure the integration function. Functional measurement use factorial experiments, selecting some attributes of a stimulus and combining them in factorial plans. Usually, subject's evaluations for each cell of experimental design are reported on a category scale, and each subject replicates each evaluation for more trials. Starting from subject's evaluations, functional measurement aims to quantify the value of each level of factors and its importance in the global judgment, for each subject evaluator or group of subjects.

Anderson's theory suggests that the most widely used integration rules are of three fundamental and simple kinds: additive, multiplicative and weighted average. Statistical techniques as the analysis of variance can be used to detect the integration rule on the basis of the goodness of fit.

The averaging rule in particular can account for interaction effects between factors, splitting evaluation in two components: scale value and weight, which can be identified and measured separately (Zalinski and Anderson, 1989). If scale value represents the location of the level of attribute on the response scale, the weight represents his importance into global judgment. Averaging model provides a useful way to manage interaction between factors, surpassing the assumption of independence on which most applications of multi-attribute choice models are based. However, the model presents some critical points about the estimation issue, and for this motivation its potential is not fully exploited up until now. In this research work, a new method for parameter estimation for averaging model is proposed. The method provides a way to select the best set of parameters to fit data, and aims to overcome some problems that have limited the use of the model.

According to this new method, named R-Average (Vidotto and Vicentini, 2007; Vidotto, Massidda and Noventa, 2010), the choice of optimal model is made according to so-called "principle of parsimony": the best model is the "simplest" one which found the best compromise between explanation of phenomenon (explained variance) and structural complexity (number of different weight parameters). Selection process use in combination two goodness-of-fit indexes: Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC; Schwarz, 1978). Both indexes are derived star-

ting from the logarithm of the residual variance weighted for the number of observations, and by penalizing the models with additional parameters. AIC and BIC differ in penalty function - since the BIC imposes a larger penalty for complex models than the AIC does - and are very useful for model comparison.

In this research work, two version of R-Average method are presented. This two versions are one evolution of the other, and both methods are structured in some procedures to perform estimation. Basically, R-Average consists of three procedures: EAM Procedure, DAM Procedure and Information Criteria (IC) Procedure. EAM, DAM and IC differ in constraints imposed on weights during the optimization process.

EAM Procedure constrains all the weight within each factor to be equal, estimating an Equal-weight Averaging Model. This model is the optimum in terms of parsimony, because it presents the smallest number of parameters (one single weight for all levels of each factor). In fact, it is defined as “parsimonious” a simple model, in which the weights are equal. Differently, DAM Procedure does not impose constraints on the weights, leaving their free to vary. Thus, this procedure may converge to a complete Differential-weight Averaging Model, which is the less parsimonious model (i.e. all the weights of each level of each factor are different).

The core of R-Average method is the Information Criteria Procedure. This procedure is based on idea that, from a psychological point of view, a simple model is more plausible than a complex model. For this reason, estimation algorithm is not oriented to search parameters that explain the greater proportion of variance, but search a compromise between explained variance and model complexity. A complex model will be evaluated as better than a simpler one only if the allows a significantly higher degree of explanation of phenomenon.

IC Procedure search the model, trying to keep (in the “forward” version) or to make (in the “backward” version) all the weights equal. In the forward version, the procedure starts from the EAM model and spans all the possible combination of weights, modifying it: initially one by one, then two by two, then three by three and so on. For each combination, the procedure tries

to diversifies weights. From time to time, using BIC and AIC indexes, the procedure selects the best set of parameters and assume the selected model as reference for the following step (if an evidence of improvement is found). In the backward version, the procedure starts from the DAM model and spans all the possible combinations of weights, trying to equalize them. BIC and AIC are used to compare the new models with the reference model: if a new model is detected as better than the reference one, it will used as new reference for following steps.

Finally, all the estimated models by the procedures are compared, and the best model based on information criteria is preferred.

The original formulation of the averaging model was modified in the evolution of the basic R-Average method. This reformulation considers the weight not as simply w parameters but as $w = exp(t)$. This exponential transformation leads to a solution for classical problem of uniqueness which affect averaging formulation (Vidotto, 2011). Furthermore, this reformulation justifies the application of cluster analysis algorithms on weight values, necessary for the clustering procedure of experimental subjects on the basis of their similarity. In fact, the distance between two t values can be evaluated in terms of simply difference. Differently, the distance between two w values can be evaluated only in terms of ratio between them. This allows to use clustering algorithms of subjects based on matrices of proximity between parameters.

The performance of R-Average was tested using Monte Carlo studies and practical applications in three different research fields: in marketing, in economic decision theory and in interpersonal trust.

Results of Monte Carlo studies show a good capability of the method to identify parameters of averaging model. Scale parameters are in general well estimated. Differently, weight estimation is a bit more critical. Punctual estimation of the real value of weights are not precise as the estimation of scale values, in particular as the standard deviation of the error component in observed data increases. However, estimations appears reliable, and equalities between weights are identified. The increasing of the number of experimental trials can help model selection when the errors present a greater standard

deviation. In summary, R-Average appear as an useful instrument to select the best model within the family of averaging models, allowing to manage particular multi-attribute conditions in functional measurement experiments.

R-Average method was applied in a first study in marketing field. In buying a product, people express a preference for particular products: understanding cognitive processes underlying the formulation of consumers' preferences is an important issue. The study was conducted in collaboration with a local pasta manufacturer, the Sgambaro company. The aims of research were three: understand the consumer's judgment formulation about a market product, test the R-Average method in real conditions, and provide to Sgambaro company useful information for a good marketing of its product. Two factors was manipulated: the packaging of the Sgambaro's pasta (Box with window, Box without window and Plastic bag) and the price (0.89 euro, 0.99 euro, 1.09 euro).

Analyses started considering evaluations of the product express by participants: for each subject, parameters of averaging model was estimated. Since the consumers population is presumably not homogeneous in preferences, the overall sample has been split in three clusters (simply named A, B and C) by an cluster analysis algorithm. For both Price and Packaging factors, different clusters showed different ratings. Cluster A express judgments that are positioned on the center of scale, suggesting as participants are not particularly attracted by this products. By contrast, Cluster B express positive judgments, and Cluster C express globally negative with the exception of the package "box with window". For packaging, it observes that the box with window, although is not the preferred one in the three clusters, has always positive evaluations, while judgments on other packaging are inconsistent across groups. Therefore, if the target of potential consumers for the product is the general population, the box with window can be considered the most appreciated packaging.

Moreover, in Cluster C ANOVA shows a significant interaction between Price and Packaging. In fact, estimated parameters of averaging model show that Cluster C is greatly affected by a high price. In this cluster the highest price had a double weight in the final ratings, therefore the positive influence

on the judgment of the “box with window” packaging could be invalidated. It’s important to notice that the group which is more sensitive to the high price is also that one which gave the lowest ratings compared to the other clusters.

In a second experiment, the R-Average method has been applied in a study in the field of economic decision making under risk. The assumption that moved the study is that, when a person must evaluate an economic bet in a risky situation, person integrates cognitively the economic value of bet and the probability to win. In the past, Shanteau (1974) shown that integration between value and probability is made according a multiplicative rule. The study, as Lynch (1979), highlighted that when the situation concern two simultaneous bets, each one composed from a value and a probability, judgments for double bet is different to the sum of judgments for single bets. This observation, named subadditivity effect, violate the assumptions of Expected Utility Theory.

The proposed study analyze the convenience/satisfaction associated with single and duplex bets. The study proposed to participants two kind of bets. A first group of bets involved a good (Mobile Phones), and the other one, a service (free SMS per day); to each good/service was associated the a probability to obtained him. Two experimental conditions was defined. In the first condition, subjects judge bets considering that phones come from a good company, and SMS service came from a untrustworthy provider. In the reverse condition, subjects judge bets considering that phones was made with low-quality and come from a untrustworthy company, and SMS service come from a strong and trustworthy provider.

For duplex bets, the presence of averaging integration model was hypothesized, and the parameters of model was estimated using R-Average on each subject. Results show that the integration in presence of a duplex bet is fully compatible with an averaging model: the averaging and not adding appear the correct integration rule.

In the last experiment, averaging model and R-Average methodology were applied to study trust beliefs in three contexts of everyday life: interpersonal, institutional and organizational. Trusting beliefs are a solid persuasion that

trustee has favorable attributes to induce trusting intentions. Trusting beliefs are relevant factors in making an individual to consider another individual as trustworthy. They modulate the extent to which a trustor feels confident in believing that a trustee is trustworthy. According to McKnight, Cummings and Chervany (1998), the most cited trusting beliefs are: benevolence, competence, honesty/integrity and predictability.

The basic idea under the proposed study is that beliefs might be cognitive integrated in the concept of trustworthiness with some weighting processes. The R-Average method was used to identify parameters of averaging model for each participant. As main result, analysis shown that, according to McKnight et al. (1998), the four main beliefs play a fundamental role in judging trust. Moreover, agreeing with information integration theory and functional measurement, an averaging model seems to explain individual responses. The great majority of participants could be referred to the differential-weight case. While scale values show a neat linear trend with higher slopes for honesty and competence, weights show differences with higher mean values, still, for honesty and competence. These results are coherent with the idea that different attributes play a different role in the final judgment: indeed, honesty and competence seem to play the major role while predictability seems less relevant. Another interesting conclusion refers to the high weight of the low level of honesty; it seems to show how a belief related to low integrity play the most important role for a final negative judgment. Finally, the different tilt of the trend for the levels of the attributes in the three situational contexts suggests a prominent role of the honesty in the interpersonal scenarios and of the competence in the institutional scenarios.

In conclusion, information integration theory and functional measurement seem to represent an interesting approach to comprehend the human judgment formulation. This research work proposes a new method to estimate parameters of averaging models. The method shows a good capability to identify parameters and opens new scenarios in information integration theory, providing a good instrument to understand more in detail the averaging integration of attributes.

Capitolo 1

La Teoria dell'Integrazione delle Informazioni

1.1 Introduzione

I processi di presa di decisione sono parte fondamentale della vita quotidiana. Nel corso della seconda metà del novecento, la descrizione di questi processi attraverso l'uso di modelli matematici ha interessato moltissimi studiosi: le cosiddette teorie “multi-attributo” hanno giocato un importantissimo ruolo nella descrizione dei meccanismi sottostanti la presa di decisioni, grazie alla formulazione di sofisticati modelli previsionali delle scelte operate dalle persone in molteplici situazioni.

In generale, i modelli di “scelta multi-attributo” poggiano su un substrato teorico che assume che il giudizio su un particolare stimolo ambientale o una decisione in una specifica situazione derivi da un processo a tre stadi (Anderson, 1981; Lynch, 1985): la valutazione di ogni singola componente dello stimolo, l'integrazione di questi valori in una risposta interna e infine l'espressione di questa risposta interna in termini di giudizio (e conseguente scelta).

Questa concezione ha trovato moltissime applicazioni in svariati settori della psicologia. Inoltre, grazie all'affinamento di appropriate tecniche sperimentali, gli psicologi hanno potuto mettere alla prova e testare la verosimi-

glianza di molti modelli multi-attributo, alla ricerca del migliore adattamento di questi alle specifiche situazioni proposte dalla vita quotidiana.

La Teoria dell'Integrazione delle Informazioni (IIT: Anderson, 1981, 1982) si è occupata di descrivere i processi psicologici sottostanti la presa di decisione, e ha proposto un generico framework applicabile in molteplici contesti. La teoria propone un substrato teorico (Algebra Cognitiva) che viene integrato da una parte metodologica (Misurazione Funzionale) necessaria alla valutazione dell'adattamento ai contesti reali dei modelli proposti. La IIT ha trovato nel corso degli anni molteplici applicazioni a livello sperimentale (si veda ad es. Anderson, 1991). Tali studi sono stati volti a individuare le modalità con cui più stimoli vengono integrati in un'unitaria rappresentazione mentale e a quantificare l'apporto di ogni componente al concetto unitario formulato. Le pratiche sperimentali prevedono di sottoporre alla valutazione di un campione di soggetti una serie di alternative formulazioni di uno stesso stimolo (profili), generate combinando fra loro le modalità o intensità che ogni componente che descrive lo stimolo può assumere. Ogni soggetto sperimentale esamina lo stimolo nella sua globalità, per fornire una risposta.

Di seguito verranno approfonditi nel dettaglio tutti questi aspetti teorici e metodologici della IIT: saranno descritti i concetti fondanti la teoria, i modelli d'integrazione e le pratiche sperimentali in uso per l'identificazione dei modelli. Un modello in particolare, quello a media ponderata, sarà analizzato e discusso nei suoi punti di forza e nelle sue criticità. Proprio tali criticità, che motivano il presente lavoro, costituiranno lo spunto da cui partiranno i successivi capitoli, dove una nuova metodologia per la selezione dei modelli dell'algebra cognitiva e la stima dei parametri dei modelli a media ponderata sarà presentata e applicata.

1.2 Il processo di formulazione del giudizio

La Teoria dell'Integrazione delle Informazioni concepisce i processi cognitivi che portano all'integrazione di più informazioni in un unico concetto come suddivisi in tre fasi, ognuna regolata da una specifica funzione:

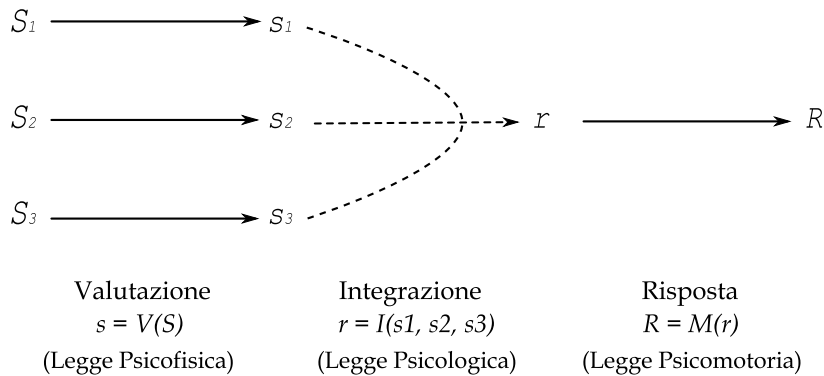


Figura 1.1: Diagramma d'integrazione. Gli stimoli percepiti S vengono trasformati in stimoli impliciti s e integrati in r , la risposta implicita, dalla quale deriva a sue volta R , la risposta esplicita.

- Funzione di Valutazione: $V(S) = s$
- Funzione di Integrazione: $I(s) = r$
- Funzione di Risposta: $M(r) = R$

Il processo è quindi realizzato in tre distinti stadi sequenziali. Alle singole caratteristiche fisiche S di uno stimolo viene attribuito un valore implicito s , e tutti i valori impliciti vengono integrati in un unico concetto unitario r , riassuntivo di tutte le caratteristiche considerate. Infine, una trasformazione strettamente monotona M converte la risposta implicita in risposta esplicita R , espressa in termini di giudizio su una particolare scala di risposta.

Per ciò che concerne la funzione di valutazione, si assume che, a livello inconsapevole, ogni soggetto utilizzi una rappresentazione numerica astratta dei valori di ogni singola componente. La IIT, in particolare, assume che ogni componente venga valutata indipendentemente dalle altre. Comunemente, ogni parte costitutiva dello stimolo sottoposto a valutazione viene denominata *attributo*.

Questo processo di valutazione porta all'assegnazione di un valore implicito alle singole parti costituenti lo stimolo. Successivamente avviene un'integrazione di questi valori, che conduce alla formulazione di un giudizio complessivo. È a questo livello che entrano in gioco i modelli di scelta, che descrivono le modalità attraverso le quali viene realizzata l'operazione d'integrazione. L'Algebra Cognitiva propone infatti tre modelli algebrici che

regolano la funzione di valutazione: il modello additivo (Anderson, 1962), il modello moltiplicativo (Anderson, 1970; Anderson e Shanteau, 1970) e il modello a media ponderata (Anderson, 1965, 1967), anche detto *averaging*.

Nelle tipiche ricerche sperimentali, partendo dalle risposte globali emesse in presenza di uno stimolo descritto da più attributi, si cerca di risalire a posteriori agli originali valori degli attributi che hanno contribuito a comporre la risposta finale. Questo lavoro viene realizzato servendosi di tali modelli algebrici, che proprio per questo loro particolare uso vengono definiti *decompositivi* (Green e Srinivasan, 1978): una risposta unitaria viene scorporata nei valori dei singoli attributi che l'hanno generata. Tale possibilità si basa sull'idea di poter misurare congiuntamente l'effetto di due fattori (Luce e Tukey, 1964; Green e Rao, 1971; Krantz e Tversky, 1971), e gran parte della teoria relativa ai modelli multi-attributo si incentra proprio sui metodi per la decomposizione delle risposte globali nei singoli valori degli attributi che le hanno generate¹.

In ultima istanza, la teoria propone che dal concetto unitario formulato a seguito dell'integrazione, chiamato risposta implicita, derivi direttamente la risposta esplicita, che viene espressa dal soggetto valutatore su una particolare scala di misura. Usualmente, nell'ambito della IIT, si fa riferimento a scale numeriche di tipo continuo (Anderson, 1962, 1970). La teoria presuppone che tra risposta implicita, indicata con r , e risposta manifesta, indicata con R , ci sia una relazione lineare. Vale quindi la regola:

$$R = C_0 + C_1 r \quad (1.1)$$

dove C_0 e C_1 sono due costanti che assegnano alla risposta implicita rispettivamente lo zero e l'unità della scala di misura, che non hanno bisogno di essere conosciuti in quanto sono automaticamente assorbiti nei valori dello stimolo (Anderson, 1981, p. 15; Anderson, 1982, pag. 58). Dunque, dal concetto di risposta implicita deriva direttamente quello di risposta manifesta,

¹Un elenco delle pubblicazioni presenti in materia sarebbe troppo lungo e oltre gli scopi del presente lavoro; tuttavia, una buona rassegna delle principali metodologie in uso è stata in passato presentata da Lynch (1985) e, più recentemente, da Carroll e Green (1995) e da Moskowitz e Silcher (2006).

che altro non è che la stessa risposta interna esplicitata su una determinata scala di misura. In questo senso, la risposta implicita in quanto tale non è misurabile, ed è conoscibile solo nella sua manifestazione estrinseca, rilevabile e quantificabile. La scala permette di dare una metrica non solo alla risposta finale ma anche alle componenti che l'hanno determinata, che possono così essere quantificate.

1.3 L'Algebra Cognitiva

Le valutazioni numeriche implicite degli attributi di uno stimolo sono usualmente chiamati dalla IIT “valori di scala”, in quanto vengono misurati sulla medesima scala su cui è espressa la risposta unitaria finale R . Nei tre modelli proposti dall'algebra cognitiva, questi valori sono identificati dai parametri s , che possono quindi essere concepiti in termini di *location*, cioè come posizionamento del valore sulla scala di risposta.

La teoria prevede che ai valori di scala siano associati dei pesi. I pesi (*weights*) rappresentano l'importanza che il particolare attributo assume nel giudizio globale, e nei modelli sono indicati dal parametro w . Come vedremo più avanti, in realtà solo nel modello averaging questi parametri assumono rilevanza.

Quindi, nella formulazione più generica, la IIT propone che la risposta r sia funzione dei valori di scala e dei pesi assunti da tutti gli attributi che descrivono lo stimolo. Nel caso in cui l'integrazione avvenga fra due attributi, che chiameremo A e B , i quali rispettivamente possono assumere le modalità i e j , la regola generale d'integrazione sarà:

$$r_{A_i B_j} = \mathbb{I}(s_{A_i}, w_{A_i}; s_{B_j}, w_{B_j}) \quad (1.2)$$

La funzione d'integrazione \mathbb{I} definisce il valore della risposta r . Come anticipato, l'algebra cognitiva propone tre fondamentali forme che tale funzione \mathbb{I} può assumere: additiva, moltiplicativa e averaging. Nei successivi paragrafi verranno approfonditi i tre modelli proposti dalla IIT, incentrandosi parti-

colarmente su quello averaging, principale tematica del presente lavoro di ricerca.

1.3.1 I modelli additivo e moltiplicativo

Il modello additivo descrive l'integrazione come una somma tra i valori attribuiti alle componenti che definiscono lo stimolo. Il valore di ogni attributo è dato dal prodotto tra i valori di scala e i rispettivi pesi, e la risposta complessiva è il risultato dalla somma dei valori di tutti gli attributi. Per K attributi, con $k = 1 \dots K$, il modello è:

$$r = \sum_{k=1}^K w_k s_k \quad (1.3)$$

Per uno stimolo descritto da due attributi A e B , il modello sarà:

$$r_{A_i B_j} = w_{A_i} s_{A_i} + w_{B_j} s_{B_j} \quad (1.4)$$

Diversamente, il modello moltiplicativo descrive l'integrazione come il prodotto fra i valori attribuiti alle componenti dello stimolo:

$$r = \prod_{k=1}^K w_k s_k \quad (1.5)$$

Per il caso a due fattori, il modello sarà:

$$r_{A_i B_j} = w_{A_i} s_{A_i} \times w_{B_j} s_{B_j} \quad (1.6)$$

Nonostante la formulazione teorica preveda la divisione tra valori di scala e pesi, sia nel modello additivo che in quello moltiplicativo i due parametri non sono realmente distinguibili (Anderson, 1981, p. 20): l'effetto di ogni attributo non può essere separato in un valore di scala e in un peso. Comunemente, i pesi vengono quindi omessi dalla formulazione dei due modelli perché confusi con i valori di scala.

Come evidenziato da Noventa, Massidda e Vidotto (in revisione), la funzione additiva modella la superficie di risposta attraverso un piano (figura 1.2,

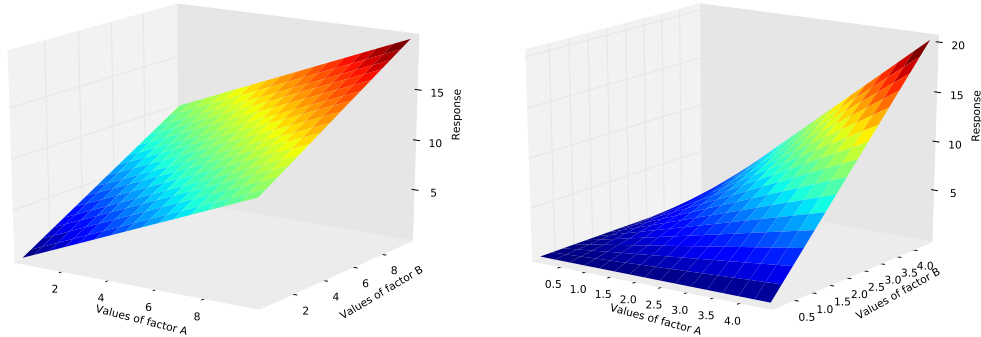


Figura 1.2: Superficie di risposta per il modello additivo (grafico di sinistra) e per il modello moltiplicativo (grafico di destra).

grafico di sinistra), mentre la funzione moltiplicativa attraverso una superficie parabolica (figura 1.2, grafico di destra).

1.3.2 Il modello averaging

Il terzo modello d'integrazione è quello che descrive la risposta come frutto di una media ponderata dei valori degli attributi. Per K attributi di uno stimolo (con $k = 0 \dots K$), la formulazione generica del modello è:

$$r = \sum_{k=0}^K \frac{w_k}{\sum_{k=0}^K w_k} s_k \quad (1.7)$$

Per il caso a due fattori, il modello può essere scritto come:

$$r_{A_i B_j} = \frac{w_0}{w_0 + w_{A_i} + w_{B_j}} s_0 + \frac{w_{A_i}}{w_0 + w_{A_i} + w_{B_j}} s_{A_i} + \frac{w_{B_j}}{w_0 + w_{A_i} + w_{B_j}} s_{B_j} \quad (1.8)$$

L'idea alla base del modello è che ogni componente che genera la risposta non possa essere considerata indipendente dalle altre. Nel modello averaging, infatti, ogni attributo che interviene nella valutazione è in grado di influenzare gli altri: le componenti sono sì valutate indipendentemente le une dalle altre, ma in fase d'integrazione si influenzano vicendevolmente. Questo fenomeno accade proprio in virtù della ponderazione: dato che i parametri di peso di

tutti gli attributi compaiono al denominatore di ogni singola componente, la valutazione del singolo attributo è influenzata dai pesi assunti anche dagli altri attributi. Bisogna quindi fare una distinzione fra il *peso assoluto* w_k e il *peso relativo* $w_k / \sum_{k=0}^K w_k$. Il peso assoluto rappresenta l'effettiva incidenza che ogni attributo ha nel determinare la risposta finale, mentre il peso relativo rappresenta l'influenza dell'attributo al netto dell'importanza di tutti gli altri attributi.

Nella formulazione del modello compare una condizione in cui $k = 0$, identificata dai parametri s_0 e w_0 , denominata *stato iniziale* (Anderson, 1959, 1965; Noventa, Massidda e Vidotto, 2010). Originariamente, lo stato iniziale era stato utilizzato da Anderson (1959) in uno studio che si serviva di presentazioni seriali di stimoli, per rappresentare l'opinione che un soggetto aveva riguardo uno stimolo già prima della sua presentazione. Successivamente (Anderson, 1965, 1967), questa condizione fu inserita anche all'interno della IIT come costante I_0 , per essere poi scissa in un valore di scala s_0 e in un peso w_0 (Anderson, 1981). In generale, lo stato iniziale rappresenta una condizione interiore, una credenza iniziale relativa allo stimolo o un atteggiamento verso la situazione sperimentale (Zalinski, 1984; Zalinski e Anderson, 1991).

In sintesi, rispetto a tanti altri modelli proposti in letteratura (anche esterni alla IIT), l'averaging permette di superare l'assunto d'indipendenza fra gli attributi, suddividendo ogni singola componente attraverso una rappresentazione a due parametri: valore e peso. Il valore complessivo di un attributo non è dato, come nei modelli additivi e moltiplicativi, dai soli parametri che si riferiscono all'attributo stesso. Nell'averaging, il valore di un attributo è dato sì dal prodotto tra il suo valore di scala e il suo peso, ma *al netto* del valore dei pesi di tutti gli altri attributi in gioco.

Le proprietà del modello

La media ponderata gode di alcune basilari proprietà, la cui trattazione è fondamentale per la piena comprensione del modello averaging. La prima (proprietà traslativa) dice che, aggiungendo (o sottraendo) a tutti i valori s una medesima quantità costante c , il risultato è incrementato (o ridotto)

della stessa quantità c . Assumendo per semplicità un modello d'integrazione a due attributi, con $s_0 = w_0 = 0$, la formula (1.8) diventa:

$$r = \frac{w_1 s_1 + w_2 s_2}{w_1 + w_2} \quad (1.9)$$

Se la proprietà appena esplicitata fosse vera, avremmo che aggiungere una quantità costante c alla funzione risposta descritta nella (1.9) sarebbe uguale ad aggiungere tale quantità c ai soli parametri s :

$$\frac{w_1 s_1 + w_2 s_2}{w_1 + w_2} + c = \frac{w_1(s_1 + c) + w_2(s_2 + c)}{w_1 + w_2} \quad (1.10)$$

Raccogliendo per c , il membro destro dell'equazione (1.10) può essere riscritto come:

$$\frac{w_1 s_1 + w_2 s_2 + c(w_1 + w_2)}{w_1 + w_2}$$

da cui, semplificando, si ricava:

$$\frac{w_1 s_1 + w_2 s_2}{w_1 + w_2} + c$$

che corrisponde al membro sinistro dell'equazione (1.10), come enunciato dalla proprietà.

La seconda proprietà dice che, moltiplicando (o dividendo) tutti i pesi w per una stessa quantità c diversa da zero, la media ponderata non cambia. In riferimento all'equazione (1.9), sarà:

$$\frac{w_1 s_1 + w_2 s_2}{w_1 + w_2} = \frac{c w_1 s_1 + c w_2 s_2}{c w_1 + c w_2} \quad (1.11)$$

Raccogliendo per c , il membro destro dell'equazione (1.11) diventa:

$$\frac{c(w_1 s_1 + w_2 s_2)}{c(w_1 + w_2)}$$

A questo punto è immediato vedere come, semplificando per c il membro di destra, l'equazione risulti un'identità.

Una terza proprietà dice che la somma dei pesi relativi è pari a 1, ovvero, facendo sempre riferimento alla (1.9):

$$\frac{w_1}{w_1 + w_2} + \frac{w_2}{w_1 + w_2} = 1 \quad (1.12)$$

È infatti immediato vedere come la somma dei due pesi relativi risulti una frazione con il medesimo valore al numeratore e al denominatore, pari quindi all'unità.

Queste proprietà hanno delle fondamentali ripercussioni sull'algebra cognitiva. La prima proprietà dimostra come uno spostamento di grandezza c dei valori s sulla scala di misura si ripercuota linearmente sulla risposta r (e conseguentemente su R), la quale verrà anch'essa spostata di una medesima grandezza c . La seconda proprietà dimostra come i pesi, che modulano l'effetto dei parametri s , non abbiano una scala di misura definita: se moltiplicati (o divisi) tutti per una stessa costante la risposta finale non verrà alterata. Quindi, i pesi possono essere misurati su una qualsiasi scala, indipendentemente dalla quella su cui sono misurate le risposte; l'unico vincolo è che i rapporti fra essi rimangano costanti nel passaggio da una scala all'altra. Per questo motivo, è possibile fissare un parametro w di riferimento e stabilire i valori dei pesi sulla base del rapporto con quello fissato, che così assumerà il valore $w = 1$.

1.4 Il concetto di peso

Rispetto ai modelli di tipo additivo e moltiplicativo, i parametri di peso nel modello averaging assumono un ruolo decisivo. Il peso è stato qui introdotto in termini di importanza di un attributo; tuttavia, nel corso dei decenni, il termine "importanza" è stato utilizzato con accezioni molto diverse fra loro, e ciò ha portato a incomprensioni e divergenze fra gli esperti, rispecchiate da una bassa validità convergente e nomologica fra le ricerche sul campo (Van Ittersum, Pennings, Wansink e van Trijp, 2007)

Fondamentalmente, l'importanza di un attributo è un concetto multidimensionale che può assumere tre diverse accezioni (Myers e Alpert, 1977):

può essere concepita in termini di *salienza*, di *rilevanza* oppure di *incidenza*. La salienza (*saliency*) è il grado di facilità attraverso cui un attributo di uno stimolo può essere richiamato in mente o riconosciuto. La rilevanza (*relevance*) si riferisce al grado in cui un attributo è in grado di fornire un beneficio che soddisfi valori e desideri considerati come rilevanti. Infine, l'incidenza (*determinance*) riflette l'importanza che un attributo ha nel determinare il giudizio e la scelta.

I metodi decompositivi, dei quali la misurazione funzionale fa parte, fanno riferimento all'ultima definizione (Van Ittersum et al., 2007). Nei modelli multi-attributo in particolare, l'importanza è rappresentata dal peso che un individuo assegna al livello dell'attributo sottoposto a giudizio, ed è proprio il peso a determinarne l'impatto nella valutazione complessiva. Dunque, nella successiva trattazione, si farà riferimento al concetto di importanza in termini di "incidenza" che un attributo ha nel determinare il giudizio.

Per ciò che concerne il modello averaging, si è visto come complessivamente i coefficienti dei valori di scala (pesi relativi) abbiano un'incidenza pari a 1 (equazione 1.12). Se i pesi assoluti di entrambi gli attributi sono uguali, allora anche i pesi relativi lo saranno, e gli attributi si spartiranno equamente l'incidenza sulla risposta complessiva.

All'aumento di un peso assoluto, però, si osserva un aumento del peso relativo associato allo stesso attributo e una diminuzione di tutti gli altri pesi relativi (perché la loro somma dev'essere pari a 1). Ciò significa che se l'incidenza di un attributo aumenta, diminuisce quella degli altri attributi. Un peso elevato estremizza il valore di un attributo, rendendolo più positivo o più negativo a seconda della polarità del valore di scala. Diversamente, un peso basso rende ininfluenza l'attributo. Di conseguenza, proprietà di uno stimolo valutate molto positivamente o molto negativamente potrebbero comunque essere tenute in considerazione solo in maniera marginale al momento dell'espressione del giudizio.

1.5 La Misurazione Funzionale

La misurazione funzionale (Anderson, 1970, 1977), che costituisce il cuore metodologico della IIT, prevede l'utilizzo di strumenti statistici per l'identificazione del modello d'integrazione a partire da un gruppo di risposte osservate R . La metodologia proposta si serve di applicazioni sperimentali, dove gli attributi di uno stimolo diventano i fattori sperimentali, e le modalità o intensità di ogni attributo diventano i livelli dei fattori. Gli esperimenti condotti nell'ambito della IIT si servono quindi di piani fattoriali. Ogni stimolo, derivato da una particolare combinazione di fattori, viene valutato da un campione di soggetti, che esprime un giudizio di tipo *rating*. Successivamente, a partire dai giudizi raccolti, per ogni soggetto si identifica la funzione d'integrazione.

Per esempio, uno studio condotto nel campo del marketing (Massidda, Noventa e Vidotto, 2009), ha sottoposto alla valutazione di un campione di 46 persone un prodotto di genere alimentare (latte fresco alta qualità). I prodotti venivano descritti secondo due caratteristiche: la marca e il prezzo, che costituivano i fattori sperimentali, e sono state scelte tre marche e tre livelli di prezzo. Dunque, sono stati ottenuti $3 \times 3 = 9$ alternativi profili dello stesso prodotto. A ogni soggetto era chiesto di valutare, su una scala da 0 a 20, la propensione all'acquisto del prodotto. Infine, a partire dalle risposte osservate, si è individuata la funzione d'integrazione utilizzata da ogni soggetto per formulare il giudizio.

La misurazione funzionale prevede la presentazione di uno stesso profilo più volte a uno stesso soggetto sperimentale (Anderson, 1982, p. 64). Questo metodo permette di controllare la variabilità insita nel giudizio umano, valutandone l'entità sulle risposte fornite per uno stesso stimolo. Avendo più risposte per ogni profilo, sarà possibile introdurre una componente di varianza per il singolo partecipante, e ciò permetterà di condurre analisi separate sui singoli soggetti.

L'approccio classico utilizzato dalla misurazione funzionale per l'individuazione del modello d'integrazione è l'analisi della varianza (ANOVA). Si assume che, per ogni soggetto, le risposte registrate in corrispondenza delle

singole celle del piano fattoriale² corrispondano, a meno di un errore casuale, alle risposte R (che come visto costituiscono delle scale lineari delle risposte interne r). Per il caso a due fattori, sarà:

$$R_{A_i B_j} = C_0 + C_1 r_{A_i B_j} + e_{A_i B_j} \quad (1.13)$$

per ogni replicazione di una stessa cella del disegno fattoriale. Si ha quindi che la risposta osservata R risulta dalla combinazione lineare della risposta implicita r , sommata a una componente d'errore e , il quale si assume distribuito come una normale a media zero. Dato che per ogni soggetto si registrano più risposte per una medesima cella del disegno sperimentale (replicazioni o trials), si possono utilizzare le medie di tutte le risposte fornite in una medesima condizione per stimare le risposte R . A questo punto, grazie all'analisi della varianza e in particolare alla valutazione degli effetti d'interazione tra i fattori, sarà possibile distinguere se l'operazione d'integrazione che ha portato alla formulazione delle risposte osservate sia stata di tipo additivo piuttosto che moltiplicativo oppure averaging.

1.6 L'identificazione del modello

1.6.1 Modelli additivi e moltiplicativi

Il teorema del parallelismo (Anderson, 1981, pp. 15-16) sostiene che, se il modello d'integrazione è additivo, il grafico delle medie marginali apparirà come un fascio di rette parallele: l'eventuale scostamento dal parallelismo osservato sarà dovuto puramente alla componente d'errore. Grazie all'ANOVA è possibile valutare se e quanto lo scostamento dal parallelismo sia dovuto a effetti casuali: se è trascurabile, si osserverà una non significatività della componente d'interazione tra i fattori. Quindi, se il modello d'integrazione è additivo, l'adattamento di un modello lineare di tipo ANOVA che non contiene il termine d'interazione risulterà migliore del modello che lo contiene.

²Dove ogni cella del piano corrisponde a uno stimolo il cui profilo è ottenuto dalla combinazione di uno dei livelli di ogni fattore sperimentale.

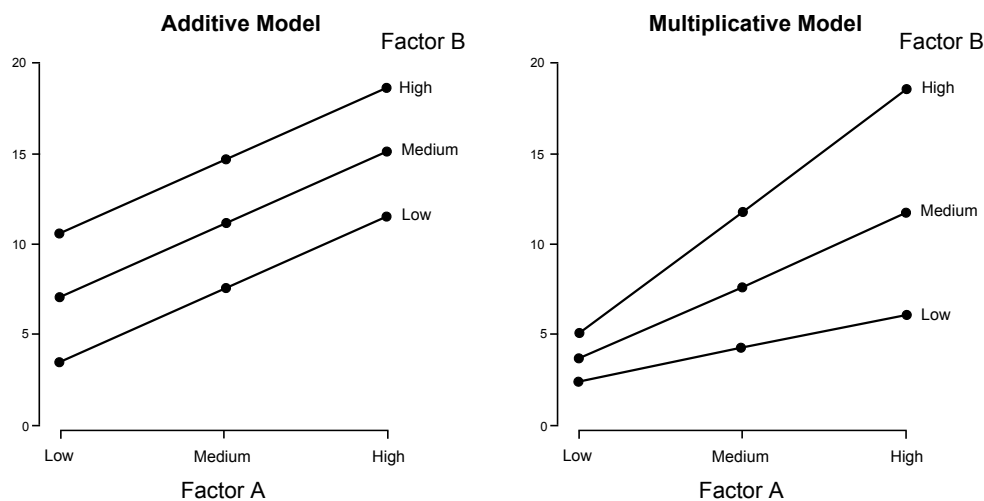


Figura 1.3: Grafico delle medie marginali per un ipotetico disegno a due fattori $A \times B$ a tre livelli ognuno. Se il pattern di parallelismo (grafico di sinistra) mette in luce l'evidenza di un modello additivo, l'andamento a ventaglio lineare (grafico di destra) evidenzia l'esistenza di un modello moltiplicativo.

Inoltre, le medie marginali delle celle del disegno fattoriale risulteranno essere scale lineari dei parametri s (Anderson, 1981, pp. 16-17).

Il teorema del ventaglio lineare (Anderson, 1981, pp. 41-42) sostiene che, se il modello d'integrazione è moltiplicativo, il grafico delle medie marginali assumerà la forma di un ventaglio lineare: la distanza fra le rette che congiungono le medie del primo fattore tenderà ad aumentare con l'aumento del valore dei livelli del secondo fattore.

L'ANOVA fornisce un metodo per distinguere il modello additivo da quello moltiplicativo: se il modello sottostante alle risposte osservate è moltiplicativo, allora, diversamente dal caso additivo, la componente d'interazione risulterà significativa. Gli effetti dei fattori possono essere scomposti utilizzando i contrasti ortogonali polinomiali; se il modello è moltiplicativo, allora l'effetto sarà tutto concentrato nell'interazione fra le componenti lineari dei fattori (Anderson, 1982, pp. 72-79). Proprio per questo motivo, gli esperimenti condotti con la metodologia della misurazione funzionale richiedono fattori misurati su almeno tre livelli: solo in questo modo sarà possibile testare l'ipotesi della linearità contro altre ipotesi alternative (interazioni delle componenti quadratiche, cubiche, ecc.).

1.6.2 Modelli averaging

Uno dei principali problemi che si presentano in misurazione funzionale riguardano la distinzione tra i modelli additivo e moltiplicativo e quello averaging. Quando il modello d'integrazione è a media ponderata, infatti, le medie marginali possono disporsi secondo pattern anche molto diversi fra loro. Questo accade a causa della natura non lineare del modello.

Fondamentalmente, nel modello averaging tutto dipende dalla configurazione dei pesi. Quando per ogni fattore i pesi sono uguali fra tutti i livelli, allora il modello sarà indistinguibile da quello additivo, e le medie marginali assumeranno una configurazione analoga a quella che si osserva in figura 1.3 (grafico di sinistra). In questo caso, nella formula (1.8) avremo che $w_{A_i} = w_{A_{i+1}}$ e $w_{B_j} = w_{B_{j+1}}$. Tale modello, che prevede pesi uguali all'interno dei fattori, viene chiamato Equal-weight Averaging Model (EAM: Wang e Yang, 1998). Nel caso i pesi assumano tale configurazione, nella formula (1.8) i pesi relativi assumeranno un valore costante per tutti i livelli dei fattori. Se indichiamo con w_A e w_B i pesi associati rispettivamente a tutti i livelli del fattore A e a tutti i livelli del fattore B , avremo che nella formula (1.8) i pesi relativi che moltiplicano i valori di scala s saranno costanti per tutte le celle del disegno sperimentale: $w_A/(w_0 + w_A + w_B)$ per il fattore A e $w_B/(w_0 + w_A + w_B)$ per il fattore B , a prescindere dal livello i assunto da A e dal livello j assunto da B . Inoltre, se i pesi sono uguali anche tra diversi fattori, ovvero $w_A = w_B$, allora il modello diventa un Complete EAM.

Diversamente, quando almeno uno dei fattori presenta un peso diverso dagli altri, si è in presenza di un Differential-weight Averaging Model (DAM: Vidotto e Vicentini, 2007). In questo caso, l'ANOVA evidenzierà un significativo effetto d'interazione, e quello presentato in figura 1.3 (grafico di destra) sarà solo uno dei possibili pattern che si potranno manifestare.

Si può quindi presentare un caso particolare in cui il modello averaging si manifesta con una configurazione delle medie marginali a ventaglio, simile a quella di un modello moltiplicativo. Questo capita quando al valore di scala più basso è associato il peso più elevato, e il valore del peso diminuisce con l'aumentare del valore di scala (Anderson e Butzin, 1974; Singh, 1991).

Quindi, quando l'incremento del valore dei parametri s è accompagnato da una diminuzione dei parametri w , il modello averaging sarà indistinguibile da quello moltiplicativo.

Dunque, la configurazione che assume la funzione averaging dipende strettamente dai pesi, e sotto certe condizioni, il pattern delle medie marginali e gli effetti risultanti dall'ANOVA possono essere totalmente sovrapponibili a quelli discussi per i modelli additivo e moltiplicativo.

Il test degli effetti opposti

Il test degli effetti opposti (Anderson, 1981) viene utilizzato per distinguere un modello additivo da uno averaging. Questo test si avvale dell'uso della metodologia dei sotto-disegni (Norman, 1976; Anderson, 1982). Tale metodologia consiste nell'associare al disegno fattoriale completo uno o più disegni fattoriali che escludono uno o più fattori alla volta. Per esempio, a un disegno fattoriale $A \times B \times C$, possono essere associati i sotto-disegni $A \times B$, $A \times C$ e $B \times C$, che combinano i fattori due alla volta. Oppure, a un disegno a due fattori $A \times B$, possono essere associati i due disegni a un fattore: A , B .

Il test degli effetti opposti risulta molto utile per distinguere un modello additivo da uno averaging a pesi uguali. Tale test consiste nel confrontare le risposte ottenute dai sotto-disegni con quelle del disegno fattoriale completo. L'idea essenziale è che, se si combinano i valori dei livelli di un attributo A con un attributo B , se la funzione d'integrazione è additiva allora il valore complessivo dello stimolo corrisponderà a un incremento costante dei valori assunti dai livelli del solo attributo A . In questo caso, disponendo sull'asse orizzontale di un grafico i livelli del fattore A e congiungendo rispettivamente i valori delle risposte ottenute dallo stimolo descritto solo da A , e i valori dello stimolo descritto dall'unione tra A e B , si otterranno delle linee parallele (figura 1.4, grafico di sinistra).

Diversamente, se la funzione d'integrazione è averaging, l'unione del valore di B con ogni livello di A non porterà a un incremento costante di tutti i livelli di A . A livello grafico, questo si rifletterà in una violazione del parallelismo (figura 1.4, grafico di destra).

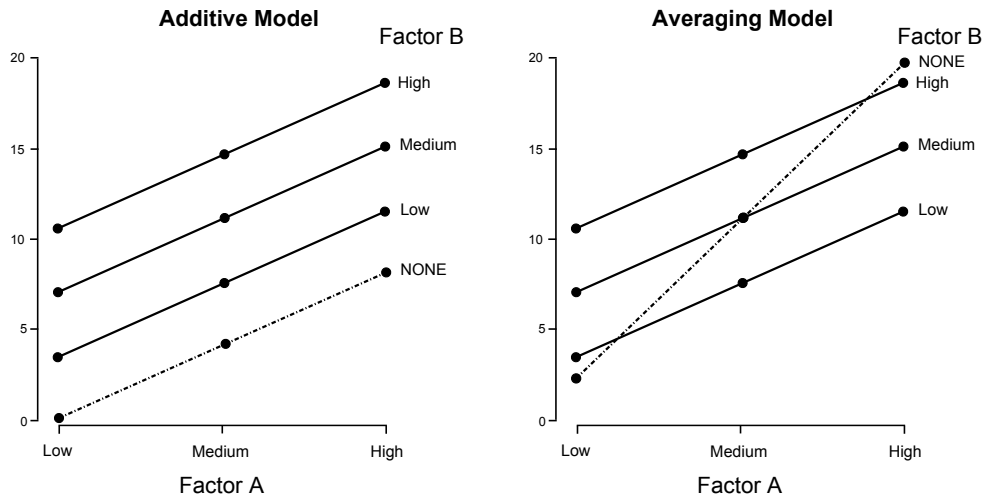


Figura 1.4: Test degli effetti opposti. Le linee continue uniscono le ipotetiche risposte relative alla combinazione dei fattori A e B ; la linea tratteggiata unisce le ipotetiche risposte relative a uno stimolo descritto dal solo attributo A per ognuno dei suoi livelli (basso, medio e alto). Nel diagramma di sinistra il parallelismo fra le rette evidenzia un'integrazione additiva, mentre nel diagramma di destra l'incrocio fra rette evidenzia un'integrazione di tipo averaging.

Il test degli effetti opposti gioca un importante ruolo anche nella distinzione tra modello moltiplicativo e averaging (Singh, 1991). Il criterio per la selezione del modello è simile a quello adottato nel caso della distinzione con l'integrazione additiva: qualora la linea ottenuta congiungendo i valori dei livelli del solo fattore A si disponesse sul grafico rispettando l'andamento a ventaglio lineare, allora l'integrazione sarà moltiplicativa; altrimenti, il modello da scegliere sarà quello averaging.

1.7 Il modello averaging a pesi diversi

Il modello averaging a pesi diversi effettua una *linearizzazione locale* di una superficie di risposta frastagliata e irregolare, che presenta picchi e depressioni (Noventa et al., in revisione). Tale linearizzazione consiste nel lisciamento della superficie in maniera diversificata per ogni cella del disegno sperimentale³. In pratica, la funzione averaging modella le risposte costruendo un

³La linearizzazione è appunto locale alla cella del disegno fattoriale, e viene realizzata in maniera diversificata per ogni incrocio dei livelli dei fattori.

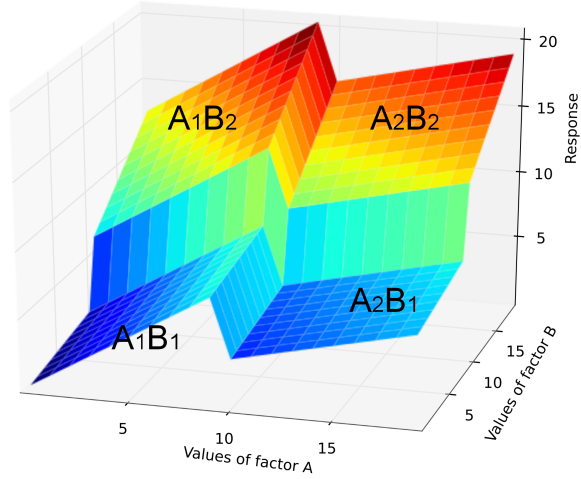


Figura 1.5: Superficie di risposta per il modello averaging in un ipotetico caso di un disegno fattoriale $A \times B$ con due livelli per fattore. Per ognuna della quattro celle del disegno sperimentale, la funzione descrive i dati attraverso uno specifico piano.

piano per ogni cella del disegno fattoriale: se il modello additivo descrive i dati con un unico piano, l'averaging li descrive attraverso l'intersecamento simultaneo di più piani (figura 1.5).

L'equazione (1.8), che descrive il modello d'integrazione a due fattori, può essere infatti riscritta sotto la forma di un modello lineare, sostituendo con β i valori dei pesi relativi:

$$r_{A_i B_j} = \beta_0 s_0 + \beta_{A_i} s_{A_i} + \beta_{B_j} s_{B_j} \quad (1.14)$$

Tuttavia, ogni parametro β non è costante ma varia a seconda di due componenti: 1) i valori assunti dai pesi degli altri fattori inclusi nel disegno sperimentale, 2) il numero di fattori inclusi nel disegno sperimentale. Sotto questo punto di vista, l'equazione (1.7), che descrive il modello averaging, non è lineare.

Per esempio, in un disegno a due fattori, con $s_0 = w_0 = 0$, il valore s_{A_i} viene moltiplicato per il coefficiente $w_{A_i}/(w_{A_i} + w_{B_j})$: la quantità risultante varia a seconda del valore assunto dal peso del j -esimo livello del fattore B . Dunque, l' i -esimo livello del fattore A assumerà il peso $w_{A_i}/(w_{A_i} + w_{B_1})$

se combinato con il primo livello di B , oppure il peso $w_{A_i}/(w_{A_i} + w_{B_2})$ se combinato con il secondo livello di B . Inoltre, lo stesso livello i del fattore A assumerà un peso ancora diverso in un disegno a tre fattori $A_i \times B_j \times C_h$, dove sarà ponderato per la quantità $w_{A_i}/(w_{A_i} + w_{B_j} + w_{C_h})$.

Quindi, il valore del peso relativo associato a ogni fattore è strettamente dipendente dai valori dei pesi assoluti degli altri fattori e dal numero di fattori inclusi nel disegno sperimentale. Ciò comporta una variazione non costante del valore della funzione averaging: in questo senso, il modello si prefigura come non lineare.

1.7.1 Identificabilità e unicità

Un'importante caratteristica che distingue il modello averaging da quello additivo e da quello moltiplicativo è il fatto che, sotto determinate condizioni, i pesi sono identificabili e possono essere misurati separatamente dai valori di scala (Zalinski e Anderson, 1989).

A titolo esemplificativo, si prenda un disegno fattoriale 2×2 : per ogni replicazione di ogni soggetto saranno osservate quattro risposte R , una per ogni cella del disegno. È possibile costruire un sistema di equazioni in cui i parametri w e s sono le incognite e le risposte R i termini noti. Assumendo le condizioni di stato iniziale $s_0 = w_0 = 0$, il sistema sarà:

$$\left\{ \begin{array}{l} \frac{w_{A_1}s_{A_1} + w_{B_1}s_{B_1}}{w_{A_1} + w_{B_1}} = R_{A_1B_1} \\ \frac{w_{A_1}s_{A_1} + w_{B_2}s_{B_2}}{w_{A_1} + w_{B_2}} = R_{A_1B_2} \\ \frac{w_{A_2}s_{A_2} + w_{B_1}s_{B_1}}{w_{A_2} + w_{B_1}} = R_{A_2B_1} \\ \frac{w_{A_2}s_{A_2} + w_{B_2}s_{B_2}}{w_{A_2} + w_{B_2}} = R_{A_2B_2} \end{array} \right. \quad (1.15)$$

Questo sistema non ha una soluzione algebrica, e risulta sotto-determinato in quanto presenta più incognite (otto) che equazioni (quattro). I parametri, quindi, non saranno identificabili. Tuttavia, il metodo dei sotto-disegni può fornire una soluzione al problema. Infatti, grazie all'uso dei sotto-disegni, è possibile ottenere un altro gruppo di risposte osservate, ottenute dalla pre-

sentazione di stimoli descritti da un unico fattore. Sempre sotto la condizione $s_0 = w_0 = 0$, ipotizzando uno stimolo descritto da un unico fattore, la risposta R per il solo fattore A sarà:

$$R_{A_i} = \frac{w_{A_i} s_{A_i}}{w_{A_i}} = s_{A_i} \quad (1.16)$$

e per il solo fattore B sarà:

$$R_{B_j} = \frac{w_{B_j} s_{B_j}}{w_{B_j}} = s_{B_j} \quad (1.17)$$

in entrambi i casi, le risposte coincidono con i valori di scala assunti dal fattore nello specifico livello. Così, i sotto-disegni possono fornire direttamente i valori di scala del sistema di equazioni (1.15), dimezzando il numero di incognite. A questo punto, per ottenere completa identificabilità dei parametri, sarà sufficiente porre un peso qualsiasi pari a 1 (par. 1.3.2) e risolvere il sistema per sostituzione.

Anche qualora l'assunzione di stato iniziale nullo non dovesse essere valida, i sotto-disegni potranno fornire ulteriori equazioni da aggiungere al sistema, che sotto determinate condizioni permetteranno di giungere all'identificabilità dei parametri (Noventa et al., 2010).

Il problema dell'unicità dei parametri del modello averaging si riferisce al fatto che, come descritto nel paragrafo 1.3.2, i pesi non possono essere identificati univocamente perché esistono più soluzioni in grado di soddisfare l'equazione del modello. La funzione averaging è infatti invariante per una trasformazione moltiplicativa dei parametri w . La scelta di fissare un peso pari a 1 è arbitraria, come è arbitraria la scelta di quale peso debba essere fissato a 1. Una soluzione a questo problema verrà discussa nel prossimo capitolo, dove una nuova metodologia per la stima dei parametri del modello sarà proposta e descritta nel dettaglio.

Capitolo 2

Il metodo R-Average

2.1 Introduzione

Nel capitolo precedente abbiamo visto come il modello averaging a pesi diversi assuma una configurazione non lineare; ciò comporta che le radici dell'equazione non possano essere esplicitate. Per risolvere sistemi di equazioni di questo tipo si è quindi costretti a ricorrere all'uso di metodi di calcolo numerico, chiamati in generale *routine di minimizzazione* o di *ottimizzazione*. Tali routine consistono in algoritmi iterativi: partendo da un valore prossimo alla radice (*starting value*) l'algoritmo va alla ricerca del minimo della funzione prestabilita. Quindi, partendo da una stima iniziale, la routine produce una successione che converge verso il valore della radice; l'algoritmo si ferma quando il punto di minimo raggiunto soddisfa il criterio di convergenza.

Alcune routine di minimizzazione sono molto sensibili ai valori di start iniziali, per cui necessitano che la successione abbia inizio da una buona approssimazione della radice. Altre routine sono meno sensibili ai valori iniziali, e altre ancora necessitano che venga esplicitato un intervallo che certamente contiene la soluzione.

La stima dei parametri per i modelli averaging richiede quindi un algoritmo iterativo che, dati dei valori iniziali, converga verso un'appropriata stima delle radici in grado di fornire una soluzione al sistema. Zalinski (1984, 1987) e Zalinski e Anderson (1991) hanno proposto il metodo AVERAGE

(implementato in Zalisnki e Anderson, 1986) per stimare i parametri di scala e di peso. AVERAGE individua i parametri che producono il migliore adattamento ai dati osservati utilizzando la routine di minimizzazione STEPIT (Chandler, 1969), applicando il criterio dei minimi quadrati. Il metodo AVERAGE cerca di minimizzare la sommatoria dei quadrati residua (RSS) del modello, definita come la sommatoria degli scarti tra valori predetti e osservati, elevati al quadrato:

$$RSS = \sum_{i=1}^N (\hat{R}_i - R_i)^2 \quad (2.1)$$

dove $i = 1 \dots N$ indicizza le risposte, \hat{R} sono le risposte teoriche e R le risposte osservate. AVERAGE stima i parametri separatamente per ogni replicazione di ogni soggetto, e prevede due procedure: una che vincola i pesi a essere uguali per tutti i livelli di un fattore, e una che permette a tutti i pesi di variare liberamente. Se la prima procedura converge verso la stima di un modello a pesi uguali (EAM), la seconda può, potenzialmente, convergere verso un caso in cui i pesi sotto tutti diversi fra loro (Complete DAM).

Uno dei problemi che si presentano nell'uso di AVERAGE è la totale assenza di una metodologia di confronto fra modelli in grado di fornire informazioni circa quale modello, EAM o DAM, debba essere considerato come migliore. Difatti, seppure un modello a pesi diversi dovesse essere in grado di spiegare i dati osservati meglio di un modello a pesi uguali, è possibile che il grado di spiegazione aggiuntiva apportato dalla ponderazione differenziale sia così esiguo da non giustificare una tale complicazione del modello (Wang e Yang, 1998).

In questo capitolo verrà presentato R-Average (Vidotto e Vicentini, 2007; Vidotto, Massidda e Noventa, 2010), un nuovo metodo per la stima dei parametri e la selezione dei modelli per il modello averaging della IIT. Tale metodo si differenzia da quello proposto in passato da Zalisnki e Anderson, sia sotto gli aspetti metodologico-procedurali che quelli applicativi. Il metodo si inserisce nell'ottica della selezione dei modelli basata su indici bayesiani e sull'applicazione del rasoio di Occam come criterio guida per la stima.

2.2 La selezione dei modelli basata sui criteri dell'informazione

Il metodo R-Average prevede che la scelta del modello migliore sia fatta in accordo col “principio di parsimonia” (rasoio di Occam): il modello migliore è quello che trova il miglior compromesso tra spiegazione del fenomeno e complessità strutturale. In termini di stima, il modello migliore sarà quello che spiega la maggior quota di varianza servendosi del minor numero possibile di parametri.

Con R-Average, la selezione del modello è realizzata per mezzo di due indici di adattamento: l'Akaike Information Criterion (AIC: Akaike, 1974) e il Bayesian Information Criterion (BIC: Schwarz, 1978). Nel caso di una stima basata sul criterio dei minimi quadrati, l'AIC può essere definito come (Burnham e Anderson, 2004):

$$AIC = n \ln\left(\frac{RSS}{n}\right) + 2k \quad (2.2)$$

mentre il BIC può essere definito come (Bogdan, Ghosh e Doerge, 2004):

$$BIC = n \ln\left(\frac{RSS}{n}\right) + k \ln(n) \quad (2.3)$$

dove n è il numero di osservazioni e k è il numero di parametri inclusi nel modello. Un valore AIC o BIC basso rispecchia un modello migliore, dove la definizione di “migliore” può essere data solo in relazione a un altro modello. Ciò comporta che, attraverso questi indici, un modello non possa essere valutato in assoluto, ma solo in relazione a un altro: i due criteri dell'informazione AIC e BIC possono essere usati per il confronto fra due modelli, ma non per valutarne l'adattamento ai dati in termini assoluti. Più elevato è l'indice, più il modello sarà penalizzato quando confrontato con un altro. Una differenza fra due indici pari a $\Delta \geq 2$ è generalmente accettata come evidenza a favore dell'esistenza di una differenza fra due modelli (Raftery, 1995). Accertata la differenza, il modello con AIC o BIC più basso sarà da preferire.

Entrambi gli indici vengo calcolati a partire dal logaritmo della RSS divisa per il numero di osservazioni. Questo valore di base è incrementato da una componente derivata da una funzione in k , che penalizza ulteriormente il modello. AIC e BIC differiscono in questa funzione di penalizzazione, dato che, rispetto all'AIC, il BIC impone una maggiore penalità ai modelli più complessi.

Come suggerito da Kuha (2004), fondamentalmente BIC e AIC differiscono nel fondamento teorico. Lo scopo dell'approccio bayesiano che muove il BIC è l'identificazione del modello che ha la più elevata probabilità di essere il *vero* modello che ha dato origine ai dati. Diversamente, l'AIC nega l'esistenza di un modello vero, e usa la capacità del modello di predire i dati come criterio per la valutazione dalla sua adeguatezza. Per queste motivazioni, il confronto fra modelli così come concepito dal metodo R-Average è basato sull'indice BIC: le differenze in AIC sono considerate solo al fine di scegliere il migliore fra due modelli quando $\Delta BIC < 2$.

Se la dimensione del campione è troppo piccola rispetto al numero di parametri considerati ($n/k < 40$), per entrambi gli indici si applica la correzione dell'errore di secondo ordine, sia per l'AIC (Sugiura, 1978; Hurvich e Tsai, 1989; Burnham e Anderson, 2004):

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1} \quad (2.4)$$

che per il BIC (Tremblay e Wallach, 2004; Almpanidis e Kotropoulos, 2008):

$$BIC_c = BIC + \frac{k \ln(n)(k+1)}{n-k-1} \quad (2.5)$$

Il metodo generale R-Average è composto da una serie di procedure algoritmiche attraverso le quali viene realizzata la stima dei parametri e viene valutato il modello. Queste procedure, basate sul metodo dei minimi quadrati, assumono una componente d'errore distribuita normalmente a media zero.

2.3 Struttura del metodo

2.3.1 Routine di minimizzazione

Il metodo R-Average si serve di una routine di minimizzazione per identificare il gruppo di parametri che minimizza la RSS. La routine scelta è BFGS (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) nella versione ottimizzata per la stima con limiti (bounding method), chiamata L-BFGS-B (Byrd, Lu, Nocedal e Zhu, 1995). Questa routine è stata scelta per la sua capacità di risolvere problemi di ottimizzazione non lineare e di lavorare con parametri vincolati a variare entro un intervallo predefinito. Infatti, come suggerito da Zalinski e Anderson (1991), stime attendibili dei parametri del modello averaging si ottengono quando alla routine di minimizzazione vengono imposti dei limiti entro i quali individuare i parametri.

L'imposizione dei limiti vincola i parametri ad assumere dei valori entro una predeterminata gamma di valori. Tale accorgimento ben si adatta alla realtà degli esperimenti nell'ambito della misurazione funzionale, dove i giudizi sono espressi su predeterminate scale di risposta che propongono un valore numerico per il giudizio minimo e uno per il giudizio massimo. Tali limiti possono costituire i vincoli entro cui i parametri s devono attenersi durante il processo di ottimizzazione. Inoltre, la possibilità di vincolare i pesi a stare entro un determinato intervallo permette di definirne la scala (par. 1.3.2).

2.3.2 Procedure per la stima e la selezione del modello

Il metodo R-Average consta di tre procedure: procedura EAM, procedura DAM e procedura Information Criteria (IC). Tutte e tre si servono della routine di minimizzazione L-BFGS-B per la stima dei parametri, ma differiscono fra loro nei vincoli imposti ai pesi durante il processo di ottimizzazione.

La procedura EAM vincola all'uguaglianza i pesi di tutti i livelli di un fattore, e conduce alla stima di un modello averaging a pesi uguali. Questo modello è il migliore in termini di parsimonia, perché presenta il minor numero di parametri. Infatti, un modello averaging è definito "parsimonioso"

quando presenta il maggior numero di pesi uguali possibile. Quando due o più pesi sono uguali vengono considerati come un unico parametro, per cui il valore k (formule 2.2-2.5) decrementa e sia l'AIC che soprattutto il BIC diminuiscono.

Diversamente, la procedura DAM non impone alcun vincolo ai pesi, lasciandoli liberi di variare. Così, questa procedura può potenzialmente condurre alla stima di un modello averaging a pesi completamente diversi, che è il meno parsimonioso (perché i pesi di tutti i livelli dei fattori sono fra loro diversi).

Il cuore del metodo è la procedura Information Criteria. Questa si fonda sull'idea che, da un punto di vista psicologico, un modello semplice è più plausibile di un modello complesso. Per questo motivo, l'algoritmo di stima non va semplicemente alla ricerca dei parametri in grado di spiegare la maggior quota di varianza possibile, ma cerca un compromesso tra quota di varianza spiegata e complessità del modello. Un modello complesso sarà valutato come migliore di uno più semplice solo se la quota di spiegazione apportata dal maggiore livello di complessità è tale da consentire un grado significativamente superiore di spiegazione del fenomeno.

Come primo passo, usando i parametri stimati dalla procedura EAM come starting values, la procedura IC stima un peso alla volta. La stima di ogni peso avviene vincolando tutti gli altri pesi a essere uguali a quelli stimati precedentemente dalla procedura EAM, il cui risultato viene considerato come modello di riferimento. In pratica, vengono stimati tanti modelli quanti sono i pesi¹. Ognuno di questi modelli viene confrontato con il modello di riferimento (EAM) servendosi degli indici BIC e AIC come esposto precedentemente (par. 2.2). La procedura IC seleziona così i parametri che producono il migliore adattamento ai dati adottando il rasoio di Occam; il modello selezionato diventerà il riferimento per il passo successivo. Se nessun modello alternativo produce un adattamento migliore ai dati rispetto a quello di riferimento, allora il modello di riferimento resterà quello EAM anche

¹Dato che potenzialmente a ogni livello di ogni fattore potrebbe essere associato un peso diverso, il numero complessivo di pesi è definito dalla somma del numero di livelli di ogni fattore. Per esempio, un disegno fattoriale 3×3 avrà $3 + 3 = 6$ pesi.

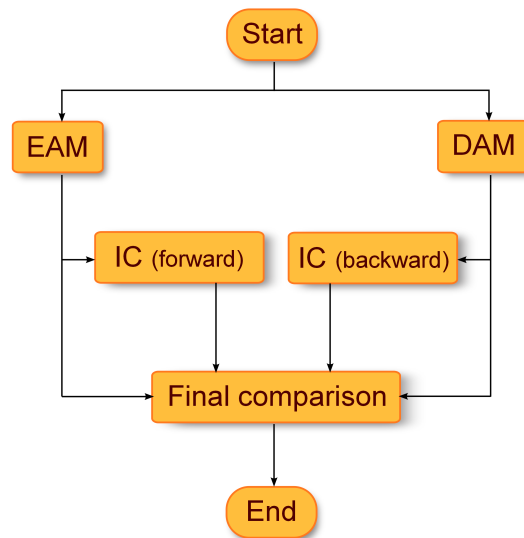


Figura 2.1: Flusso algoritmico della metodologia R-Average. Partendo da dei valori di start definiti, si stimano i modelli EAM (Equal-weight Averaging model) e DAM (Differential-weight Averaging model), e a partire da questi il modello IC (Information Criteria), nelle versioni forward e backward. Un confronto finale stabilisce il miglior risultato.

per il passaggio successivo.

Come secondo passo, la procedura modifica i pesi a coppie, cercando un ulteriore miglioramento nell’adattamento ai dati. Se viene individuata un’evidenza a favore di un miglioramento nell’adattamento, il nuovo modello sarà selezionato come nuovo riferimento per il passaggio successivo, dove i pesi saranno lasciati liberi di variare a tre a tre alla ricerca di ulteriori miglioramenti.

Così, la procedura IC testa tutte le possibili combinazioni di pesi: inizialmente lasciandoli variare a uno a uno, poi a due a due, poi a tre a tre e così via. La ricerca del modello migliore parte dal risultato della procedura EAM, che viene sostituito passo dopo passo da un nuovo modello di riferimento se gli indici BIC e AIC evidenziano un miglioramento. La procedura si ferma quando tutte le possibili combinazioni di pesi sono state testate.

Il metodo R-Average comprende due tipi di procedure IC: *forward* e *backward*. La metodologia descritta poco sopra corrisponde a un metodo di tipo forward. All’opposto, la procedura backward parte dal modello DAM, e cer-

ca il migliore adattamento ai dati vincolando i pesi all'uguaglianza: prima a coppie, poi a terne e così via. In sintesi, entrambe le versioni della procedura IC ricercano il modello in grado di spiegare la maggior quota di varianza dei dati, mantenendo, per quanto possibile, una semplicità strutturale (cioè il maggior numero di pesi uguali possibile). Nella versione forward, l'algoritmo cerca di mantenere uguali i pesi, mentre nella versione backward l'algoritmo cerca di renderli uguali.

Sia per la versione forward che per quella backward, il metodo fornisce in aggiunta una forma ridotta della procedura IC, in cui non vengono testate tutte le possibili combinazioni di pesi. Questa forma ridotta si serve di un'euristica, secondo la quale i pesi che variati singolarmente hanno prodotto un significativo miglioramento nell'adattamento ai dati, saranno quelli che con maggiore probabilità potranno produrre un ulteriore miglioramento se variati a coppie. Seguendo la stessa logica, i pesi che hanno prodotto un significativo miglioramento quando variati a coppie, saranno quelli che con maggiore probabilità produrranno un miglioramento quando variati a terne, e così via per le combinazioni di grado successivo. Quindi, la forma ridotta della procedura IC seleziona i pesi da far variare (se in versione forward) o da fissare (se in versione backward) in base al fatto che questi nel passaggio precedente abbiano prodotto o meno un significativo miglioramento nell'adattamento ai dati. In questo modo, le combinazioni da testare vengono scelte *in itinere*, e il processo termina in maniera molto più rapida se è presente qualche peso che effettivamente è diverso dagli altri.

Una volta che le tre procedure EAM, DAM e IC hanno completato la stima, i tre modelli risultanti vengono comparati usando i criteri dell'informazione: il miglior modello fra questi viene scelto come risultato finale.

2.3.3 L'identificazione delle uguaglianze fra i pesi

Il calcolo degli indici AIC e BIC richiede la definizione del valore k , che rappresenta il numero di parametri inclusi nel modello. Dato che quando due pesi sono uguali vengono considerati come un unico parametro, è necessario definire un criterio in grado di sancire l'uguaglianza fra due o più pesi.

R-Average si serve di un valore soglia Δ_w : se la distanza tra due o più pesi è inferiore o uguale a questo valore predefinito, allora tali parametri saranno considerati uguali. Dato un parametro w_i , tutti i pesi che ricadono nell'intervallo $[w_i - \Delta_w, w_i + \Delta_w]$ diventeranno un unico parametro uguale al loro valore medio.

Da quanto esposto consegue che due pesi che di per loro sarebbero diversi, ma che risultano entrambi assimilabili a un terzo peso, saranno comunque eguagliati. In questo modo, la metodologia cerca di contenere il numero di parametri del modello rispettando il principio di parsimonia.

2.4 Applicazione del metodo

Perché possa essere utilizzato, il metodo sopra descritto necessita che alcune disposizioni algoritmiche vengano definite a priori.

In primo luogo, l'uso di una routine di minimizzazione richiede dei valori di partenza per i parametri e in alcuni casi dei limiti entro cui questi devono essere compresi. Generalmente, tali valori possono essere stabiliti a partire dalla gamma della scala di risposta: il minimo e il massimo della scala potranno corrispondere al minimo e al massimo dei parametri s , e il valore di centro scala potrà costituire il valore di partenza. Nel caso in cui si ipotizzi che alcuni valori di scala possano essere molto vicini ai limiti imposti ai parametri durante il processo di minimizzazione, potrebbe risultare utile vincolare la stima in modo che i parametri vengano ricercati entro margini più ampi rispetto a quelli definiti dalla scala di risposta. Ciò eviterebbe l'effetto di compressione verso il limite (Vidotto et al., 2010) e permettere alla funzione di ottimizzazione di oscillarvi intorno simmetricamente.

Per quanto riguarda i pesi, dato che la loro scala è arbitraria e ciò che conta non sono i loro valori in assoluto quanto i rapporti che intercorrono fra essi, i limiti possono essere imposti arbitrariamente. Imponendo un valore di partenza pari a 1 per tutti i pesi, una scala che oscilla tra un valore prossimo a 0 e 10 costituisce una gamma abbastanza ampia perché i parame-

tri siano adeguatamente stimati². Particolare attenzione va posta al limite minimo, che in nessun caso può raggiungere il valore 0 in quanto la funzione obiettivo assumerebbe valore infinito. In questo caso, il valore dovrà essere comunque prossimo a zero, ma il grado di approssimazione dipenderà dall'arrotondamento decimale considerato nella valutazione qualitativa dei parametri.

In secondo luogo, il metodo richiede di specificare quale versione della procedura Information Criteria utilizzare. La versione completa è sicuramente quella da preferire, ma la versione ridotta può comunque risultare un ottimo compromesso tra bontà delle stime e vincoli imposti dalle capacità di calcolo del supporto hardware. Infatti, a seconda della dimensione del disegno sperimentale, la versione completa potrebbe richiedere un tempo di calcolo eccessivo, e in questi casi la versione ridotta risulterebbe più efficiente.

In terzo luogo, il metodo richiede di definire la minima differenza di BIC e di AIC perché due modelli siano considerati diversi. Questi due valori dovrebbero essere scelti in base al grado di conservatività che si vuole adottare nel processo di selezione; in linea generale, il valore $\Delta BIC = \Delta AIC = 2$ indicato da Raftery (1995) può essere considerato una soglia adeguata.

Infine, è necessario specificare il valore Δ_w , che dovrebbe essere scelto sulla base della scala dei pesi adottata e sulla base di assunzioni teoriche.

2.5 Differenze fra R-Average e AVERAGE

Il metodo R-Average presenta differenze rilevanti con AVERAGE. La più importante è l'utilizzo di algoritmi di selezione dei modelli basati sul principio di parsimonia e sui criteri dell'informazione, metodologia completamente ignorata in AVERAGE. Difatti, il metodo proposto da Zalinski e Anderson include due procedure per la stima di due modelli, EAM e DAM, e qui si ferma. Diversamente, il metodo R-Average prevede l'uso di una terza procedura che opera in accordo al principio di parsimonia, alla ricerca di un compromesso tra spiegazione del fenomeno e complessità del modello, e sempre in

²Nel caso una prima stima dovesse evidenziare l'inadeguatezza di questi valori, i limiti potranno essere opportunamente variati per una seconda stima.

accordo con questo principio effettua una comparazione fra i tre modelli identificati. Inoltre, come vedremo tra breve (par. 2.7), altre procedure di stima andranno a completare il metodo R-Average.

Ancora, il metodo di stima R-Average non prevede alcuna distinzione tra repliche sul singolo soggetto, risposte di più soggetti e gruppi di soggetti. La stima viene realizzata nello stesso identico modo a prescindere da come i dati sono stati raccolti: l'appropriata metodologia di analisi può essere scelta secondo assunzioni teoriche senza alcun vincolo imposto dal metodo.

2.6 Verifica dell'attendibilità: studio Monte Carlo

In passato, Vidotto e Vicentini (2007) hanno realizzato uno studio Monte Carlo per comparare la performance del metodo R-Average con quella del metodo AVERAGE. In generale, R-Average ha mostrato di fornire stime più attendibili, con una componente residua inferiore e con migliori indici di adattamento; tutti i parametri vengono stimati molto vicini al valore reale (l'errore di misura è stimato entro l'1%).

Per completare il test sull'attendibilità del metodo, successivamente sono stati realizzati altri studi di simulazione volti ad analizzarne la performance sotto determinate condizioni di stima. Le simulazioni hanno preso in considerazione il caso di un disegno fattoriale 3×3 , che costituisce il disegno base con cui viene costruita una gran parte degli studi di misurazione funzionale.

Così, un altro studio (Vidotto et al., 2010) è stato realizzato in fase di sviluppo del metodo, con lo scopo d'individuare gli eventuali punti critici per la stima e di verificare il comportamento delle procedure al variare di alcune strategie procedurali come i valori ΔBIC e ΔAIC , il valore Δ_w e la versione della procedura IC applicata (se completa oppure ridotta).

I risultati hanno evidenziato che l'attendibilità delle stime varia a seconda del tipo di parametro considerato. Per quanto concerne i valori di scala, le stime seguono una distribuzione normale centrata sul vero valore del parametro; la deviazione standard aumenta con l'aumentare della deviazione standard

degli errori, e decresce quando le stime vengono effettuate su un consistente numero di dati (ovvero un elevato numero di replicazioni di uno stesso stimolo). All'opposto, le stime dei pesi non seguono una distribuzione normale, e l'attendibilità decrementa con l'aumento della variabilità dell'errore sui dati. In ogni caso, la manipolazione di alcuni accorgimenti procedurali, come il valore Δ_w , possono migliorare la stima. Lo studio ha anche mostrato la bontà delle stime operate dalla versione ridotta della procedura IC; tuttavia, la versione completa resta molto utile in alcuni casi critici, permettendo di migliorare soprattutto l'identificazione dei pesi.

Come anticipato, questo studio Monte Carlo è stato realizzato in fase di messa a punto del metodo. Per questo motivo, si è scelto di realizzarne un secondo che possa considerarsi conclusivo della prima parte di sviluppo della metodologia e che ne rispecchi appieno l'affidabilità. Come il precedente, anche questo secondo studio si propone di verificare la performance di R-Average, analizzando però dati di natura diversificata, casi in cui si presentano particolari configurazioni di pesi e condizioni di stato iniziale non nulle. Questo nuovo studio propone quindi un'analisi più ad ampio spettro rispetto ai precedenti, e valuta l'effettiva capacità del metodo di convergere verso la selezione di un adeguato modello in grado di descrivere i dati osservati.

Le simulazioni Monte Carlo sono state realizzate nell'ambiente statistico R (R Development Core Team, 2011), utilizzando il pacchetto *rAverage* versione 0.3-5 (Vidotto, Noventa, Massidda e Vicentini, 2011), implementazione del metodo R-Average. Tale implementazione non comprende la versione backward della procedura IC, per cui questa non è stata considerata negli studi di simulazione.

2.6.1 Metodo

Partendo da un definito gruppo di parametri, venivano composte tutte le risposte teoriche \hat{R} del disegno fattoriale 3×3 attraverso la funzione *avraging*. Le risposte \hat{R} venivano generate per un definito numero di replicazioni dello stesso stimolo sul soggetto: 5, 10 oppure 15 replicazioni. Alle risposte teoriche veniva aggiunta una componente d'errore casuale distribuita normalmente,

Group	Scale values					Weights					
	s_0	w_0	$A_1 B_1$	$A_2 B_2$	$A_3 B_3$	A_1	A_2	A_3	B_1	B_2	B_3
1	0	0	5	10	15	4	1	4	1	3	3
2	2	1	5	10	15	4	1	4	1	3	3
3	0	0	5	10	15	4	1	4	2	2	2
4	2	1	5	10	15	4	1	4	2	2	2

Tabella 2.1: Gruppi di parametri utilizzati per i quattro studi Monte Carlo. I valori di scala sono costanti per tutti e quattro i gruppi e uguali tra primo e secondo fattore. Diversamente, i pesi variano tra i primi due gruppi e gli ultimi due, con le condizioni iniziali che possono essere assenti (fissate a zero) oppure presenti.

a media zero e con tre possibili varianze d'errore: $\mathcal{N}(0, 0.5^2)$, $\mathcal{N}(0, 1.0^2)$ e $\mathcal{N}(0, 1.5^2)$. Le risposte osservate R , generate dalla somma fra le risposte teoriche \hat{R} e la componente d'errore, erano vincolate a variare tra 0 e 20 per simulare delle reali risposte misurate su una scala a 21 punti.

Così, definite 1000 iterazioni della simulazione, per ogni iterazione si avevano nove possibili condizioni di stima diverse, ottenute dalla combinazione dei tre possibili numeri di repliche degli stimoli con le tre possibili varianze d'errore sui dati.

Sono stati scelti quattro gruppi di parametri (tabella 2.1) da cui comporre le risposte teoriche \hat{R} , e per ogni gruppo è stata realizzata una simulazione. I parametri sono stati scelti stabilendo un ordine crescente dei valori scala fra i livelli dei fattori, con $s_A = s_B = \{5, 10, 15\}$ per tutti e quattro i gruppi. Ciò che variava fra i gruppi erano i valori delle condizioni iniziali e dei pesi del secondo fattore. Difatti, il primo gruppo di parametri è identico al secondo, come il terzo è identico al quarto, eccetto che per lo stato iniziale: vincolato a zero ($s_0 = w_0 = 0$) nei gruppi 1 e 3 e presente ($s_0 = 2, w_0 = 1$) nei gruppi 2 e 4. I pesi del primo fattore erano sempre fissati pari a $w_A = \{4, 1, 4\}$; in questo modo, si è voluta verificare la capacità del metodo di identificare l'uguaglianza fra due pesi *entro* fattore, ponendo le uguaglianze fra due livelli gerarchicamente lontani fra di loro (il primo e il terzo). I pesi scelti per il secondo fattore erano $w_B = \{1, 3, 3\}$ per i gruppi 1 e 2 e $w_B = \{2, 2, 2\}$ per i gruppi 3 e 4. Nel primo caso, i valori sono stati scelti nell'ottica di verificare il grado in cui il metodo è in grado d'individuare l'uguaglianza inter fattore fra w_{A_2} e w_{B_1} e l'uguaglianza intra fattore fra w_{B_2} e w_{B_3} . Nel secondo caso, si

è voluta verificare la capacità del metodo di stimare un modello semilineare (Anderson, 1982), composto da un fattore a pesi parzialmente diversi (fattore A) e un fattore a pesi uguali (fattore B).

Queste configurazioni di pesi sono state scelte in quanto rispecchiano condizioni che nella pratica potrebbero facilmente verificarsi. Difatti, la peculiarità del modello averaging è la capacità di poter considerare anche gli effetti d'interazione tra i fattori sfruttando particolari configurazioni di pesi. Per esempio, un livello con un'elevata o con una bassa intensità può risultare più importante di un livello d'intensità media e incidere maggiormente nella formulazione del giudizio complessivo (Skowronski e Carlston, 1987; Jo, Nakamoto e Nelson, 2003). Ai livelli estremi, quindi, possono essere associati pesi più elevati. Studiando le configurazioni descritte sarà possibile analizzare la capacità del metodo d'identificare la presenza di questi particolari pattern.

Quindi, la dimensione dello studio Monte Carlo era: 4 gruppi di parametri \times 3 numeri di replicazioni sul ogni ipotetico soggetto (trials) \times 3 deviazioni standard della componente d'errore \times 1000 iterazioni. Sono stati realizzati due studi paralleli: uno per testare l'affidabilità della versione completa dell'algoritmo IC e uno per la versione ridotta.

Le analisi sono state realizzate utilizzando la routine di minimizzazione L-BFGS-B, definendo i valori 0 e 20 rispettivamente come limite minimo e limite massimo per i parametri s , e il valore di centro scala 10 come valore di partenza. I pesi sono stati vincolati a variare tra $1e-5$ e 10, e il valore di partenza scelto era $w = 1$. Per quanto riguarda le costanti di stato iniziale, sono stati utilizzati i valori 0 e $1e-5$ come starting values rispettivamente per s_0 e per w_0 ; s_0 variava tra 0 e 10, w_0 tra $1e-5$ e 2. Il Δ_w è stato scelto pari a 0.1, e i valori soglia per stabilire la differenza tra due modelli pari a $\Delta BIC = \Delta AIC = 2$.

Le simulazioni sono state realizzate in prima battuta utilizzando la versione completa della procedura IC. Successivamente, per i gruppi di parametri 2 e 4, è stata realizzata anche un'altra simulazione, utilizzando la versione ridotta.

2.6.2 Risultati

Le tabelle A.1-A.4, riportate in appendice A, mostrano i risultati delle simulazioni realizzate utilizzando la versione completa della procedura IC: per ogni parametro sono riportati la mediana e lo scarto inter-quantile. Le tabelle non riportano il valore assoluto dei parametri w ma la mediana dei loro rapporti (come visto, l'interesse è sui rapporti che intercorrono fra i pesi e non sui loro valori in senso assoluto). Quindi, le tabelle presentano i parametri w divisi per il peso che ha il valore più elevato (w_{A_1}), ponendo il peso che presenta il valore più basso a 1.

In generale, la stima è più attendibile quando la deviazione standard dell'errore è bassa; tuttavia, i parametri risultano identificati abbastanza bene anche quando $\sigma_\epsilon = 1.5$. Sia i valori di scala che i pesi appaiono ben stimati. I valori mediani degli s sono molto vicini ai valori reali che hanno originato le serie di dati, con distribuzioni di probabilità che appaiono tutte concentrate intorno al valore reale e con indici di asimmetria tutti compresi tra -0.1 e +0.1 (si veda la figura 2.2 per un caso esemplificativo).

Le uguaglianze fra i pesi sia all'interno dei fattori che tra diversi fattori sono generalmente identificate anche quando il valore del parametro non è identificato correttamente. Per esempio, per il primo gruppo di parametri i pesi risultano in generale leggermente sottostimati (tabella A.1); tuttavia, nonostante tale sottostima, le uguaglianze sono correttamente identificate. La capacità d'identificare le uguaglianze è molto marcata per i pesi del primo fattore, dove il valore di mediana di w_{A_1} è identico a quello di w_{B_1} in tutte le simulazioni.

In un caso (tabella A.4, simulazione: trials = 5, $\sigma_\epsilon = 1.5$), il metodo converge verso la selezione di un modello scorretto, preferendo per il primo fattore la struttura a pesi uguali rispetto a quella a pesi parzialmente diversi. Infatti, nel 52.1% delle iterazioni (tabella A.6), il metodo preferisce il risultato della procedura EAM a quelli delle procedure IC o DAM.

Quando presenti, i parametri di stato iniziale sono ben stimati (tabella A.5). Quando lo stato iniziale è nullo, il valore w_0 viene stimato prossimo allo

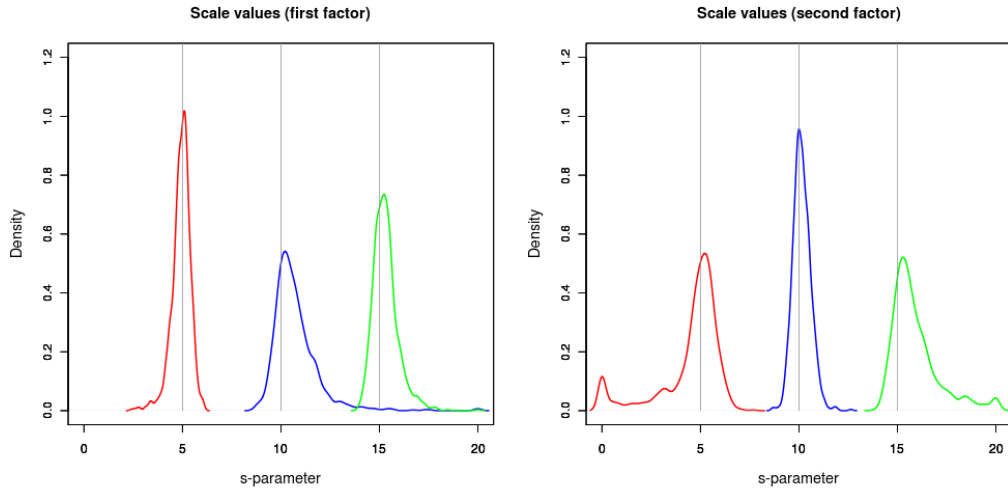


Figura 2.2: Distribuzioni dei parametri s per la simulazione realizzata sul primo gruppo di parametri per il caso 5 replicazioni e deviazione standard d’errore $\sigma_\epsilon = 1.0$. Per ogni fattore, le tre curve rappresentano le distribuzioni dei parametri di ogni livello (rosso: primo livello, blu: secondo livello, verde: terzo livello). Le linee verticali grigie rappresentano il valore reale del parametro.

zero³ nella maggior parte delle simulazioni (tabella A.5): quando $\sigma_\epsilon = 0.5$, in oltre il 90% delle simulazioni il parametro è individuato come minore di 0.1; questa percentuale diminuisce con l’aumento della deviazione standard d’errore, ma rimane comunque sopra il 60%.

Quando le stime vengono effettuate servendosi della versione ridotta della procedura IC, il metodo fornisce risultati analoghi a quelli discussi per la versione completa (tabelle A.7 e A.8).

2.6.3 Discussione

I risultati dello studio Monte Carlo hanno mostrato che, se i dati sono compatibili con un modello averaging, il metodo R-Average evidenzia una buona capacità di stima dei parametri. In linea generale, si osserva che l’aumento del numero di replicazioni della stimolazione sul singolo soggetto può aiutare la selezione del modello quando la componente d’errore presenta un’elevata varianza d’errore.

³In questi casi il parametro s_0 perde importanza, in quanto la componente al numeratore $s_0 \times w_0$ risulterà in un valore anch’esso prossimo allo zero.

I risultati necessitano di differenti considerazioni a seconda del tipo di parametro considerato. Come già evidenziato in altri studi (Vidotto et al., 2010), i parametri s sono stimati molto bene. Diversamente, la stima dei pesi risulta un po' più critica. La stima puntuale del valore dei parametri w non è precisa come per i valori di scala, soprattutto al crescere della varianza d'errore. Tuttavia, i rapporti d'ordine fra i pesi appaiono rispettati e le uguaglianze fra i parametri sono ben identificate.

In un caso (tabella A.4, simulazione: trials = 5, $\sigma_\epsilon = 1.5$), il processo di selezione converge erroneamente verso un modello a pesi uguali, e di conseguenza i pesi non sono ben identificati. Questo è probabilmente dovuto al fatto che questa simulazione è la più critica, perché presenta il più piccolo numero di replicazioni (quindi minor disponibilità di dati) e la più elevata varianza d'errore. È curioso che il problema si verifichi solo per un unico gruppo di parametri, in presenza di condizioni iniziali: ulteriori e dettagliate analisi sono richieste.

La procedura di stima IC, il cui algoritmo è basato sul principio di parsimonia, è quella che più spesso converge verso l'identificazione del modello che la selezione finale sceglierà come migliore (tabella A.6). Difatti, nel peggiore dei casi, il modello "Information Criteria" sarà scelto nel 71.6% delle iterazioni. Il metodo di selezione basato sul principio di parsimonia, quindi, fornisce un'ottima strategia per l'identificazione del modello averaging a pesi diversi, e conduce a risultati decisamente superiori alla classica procedura DAM utilizzata anche da AVERAGE.

Se l'uso della versione completa della procedura IC permette di ottenere delle buone stime, la versione ridotta fornisce risultati altrettanto buoni. Nel caso di un disegno sperimentale particolarmente ampio, la versione ridotta può fornire buoni risultati, riducendo notevolmente il tempo di calcolo.

2.7 La riformulazione del modello

Partendo dalla metodologia descritta nei precedenti paragrafi e dalle criticità emerse nello studio di simulazione, il metodo di stima è stato modificato. Tali modifiche hanno interessato principalmente il modo in cui i pesi sono

concepiti all'interno del modello. In accordo con Vidotto (2011), i parametri di peso w possono essere riscritti nella forma:

$$w = e^t$$

Il parametro di peso viene quindi concepito come l'esponentiale di un altro parametro, t . Questa trasformazione comporta notevoli vantaggi nel processo di stima.

In primo luogo, qualsiasi valore assuma t , il parametro w sarà comunque compreso nell'intervallo $(0, +\infty]$. Ciò permette di svincolare il processo di ottimizzazione dall'uso di routine di minimizzazione che impongono limiti di variazione ai parametri.

In secondo luogo, abbiamo visto come moltiplicando i pesi per una costante c , il modello averaging rimanga invariato (par. 1.3.2); la trasformazione proposta linearizza questa relazione. Difatti, dato che dal logaritmo naturale di w si ottiene t , se moltiplichiamo w per una costante c avremo che:

$$\ln(cw) = \ln(c) + \ln(w) = c' + t \quad (2.6)$$

Quindi, la trasformazione logaritmica permette di linearizzare, rendendole additive, le relazioni moltiplicative che sussistono fra i pesi. Questa riformulazione ha una ricaduta fondamentale, in quanto fornisce una soluzione al problema dell'unicità dei parametri di peso. Il problema potrebbe essere affrontato imponendo che un valore t sia uguale a zero (o, specularmente, che un valore w sia uguale a uno), e calcolando la scala degli altri parametri sulla base di questo. Tuttavia, scegliere un peso piuttosto che un altro è arbitrario. Si può quindi imporre un vincolo al sistema, determinando che la somma dei t debba essere uguale a zero (per cui, parallelamente, il prodotto dei w sarà pari a 1). Tale imposizione può essere realizzata sottraendo a ogni valore t la media di tutti i t . Questa trasformazione è invariante per qualsiasi costante c' venga sommata (o sottratta) ai parametri. Vale quindi l'uguaglianza:

$$t_i - \frac{\sum_{i=1}^n t_i}{n} = t_i + c' - \frac{\sum_{i=1}^n (t_i + c')}{n} \quad (2.7)$$

dove $i = 1 \dots n$, con n che rappresenta il numero di pesi inclusi nel modello. La (2.7) può infatti essere riscritta nella forma:

$$t_i - \frac{\sum_{i=1}^n t_i}{n} = t_i + c' - \frac{\sum_{i=1}^n t_i}{n} - c' \quad (2.8)$$

dato che nel secondo membro della (2.8) la costante c' si annulla, l'uguaglianza è soddisfatta: lo scarto dei t dalla loro media è invariante rispetto a una costante additiva. Da notare come la normalizzazione $t_i - \bar{t}$, dove \bar{t} rappresenta la media dei valori t , corrisponda al logaritmo dei pesi w divisi per la loro media geometrica:

$$t_i - \frac{\sum_{i=1}^n t}{n} = \ln \left(\frac{w_i}{\sqrt[n]{\prod_{i=1}^n w_i}} \right) \quad (2.9)$$

Inoltre, la deviazione standard dei t può essere considerata come un indice di scostamento dal parallelismo del modello. Difatti, se i pesi di uno stesso fattore sono uguali, la loro deviazione standard sarà zero. Dato che i t sono concepiti come distanze dalla loro stessa media, all'aumentare del loro valore aumentano le differenze fra i pesi, e di conseguenza le deviazioni dal parallelismo. Quest'uso della deviazione standard come indice di scostamento dal parallelismo è legittimato dal fatto che la deviazione standard è un indice invariante per qualsiasi traslazione verticale, per cui si avrà che $sd(t + c') = sd(t)$: l'aggiunta di una costante c' ai parametri t o una variazione dell'origine della scala su cui questi sono misurati non ne modifica la deviazione standard.

La trasformazione logaritmica dei pesi proposta giustifica l'uso di una distanza Δ per valutare le uguaglianze fra i parametri (par. 2.3.3), che d'ora in avanti sarà concepita come Δ_t e non più come Δ_w . Infatti, un incremento di un valore Δ in un parametro w non produce una variazione simmetrica del valore della funzione averaging rispetto a un decremento di entità Δ dello stesso parametro w . Diversamente, un incremento Δ di un parametro t produce una variazione simmetrica del valore della funzione rispetto a un decremento Δ dello stesso parametro t .

Durante il processo di ottimizzazione, assumere $\Delta_t = 0.1$ significa assu-

mere che, perché due t possano essere considerati diversi, sarà necessario che lo scarto tra i due superi il valore 0.1, altrimenti i due t saranno eguagliati. Dalla (2.6) si ricava che:

$$t + \Delta = \ln(we^\Delta) \quad (2.10)$$

Per questo motivo, un incremento pari a $\Delta = 0.1$ in un valore t si rifletterà in un incremento pari a $e^\Delta = 1.1$ in un valore w . Quindi, un incremento di un'unità decimale in un peso w non sarà considerato sufficiente perché il parametro possa essere ritenuto effettivamente diverso da un altro.

2.8 Nuove procedure di stima

In questa seconda fase di sviluppo, la metodologia R-Average è stata ampliata con nuove procedure di stima che si collocano a monte rispetto alle procedure EAM e DAM (figura 2.3). La logica che ha mosso questa ristrutturazione è stata quella di operare un aumento progressivo e graduale della complessità del modello, lavorando prima sui valori di scala per poi concludere con i pesi.

Il primo passaggio consiste nella stima di un modello nullo, che identifica un unico valore di scala per tutti i livelli di tutti i fattori, assumendo un unico peso fisso. Per la stima di questo modello è necessario quindi identificare un valore unitario per tutti i parametri s ; usualmente, la media delle risposte R costituirà il miglior stimatore dei valori di scala, mentre il valore t rimarrà fisso a zero sino allo svincolamento dei pesi operato dalle procedure EAM e DAM.

Partendo dal modello nullo, una seconda procedura differenzia i parametri s tra i fattori; viene quindi stimato un valore di scala per ogni fattore, ma entro fattori i parametri s vengono considerati uguali per tutti i livelli. Ancora, i pesi sono tutti fissati pari a $t = 0$.

Il passaggio successivo consiste nella stima di parametri s differenziali per tutti i livelli di tutti i fattori, assumendo sempre un unico peso fisso.

Passo dopo passo, i modelli vengono confrontati col precedente attraverso l'uso degli indici BIC e AIC utilizzati in combinazione (par. 2.2). Solo alla

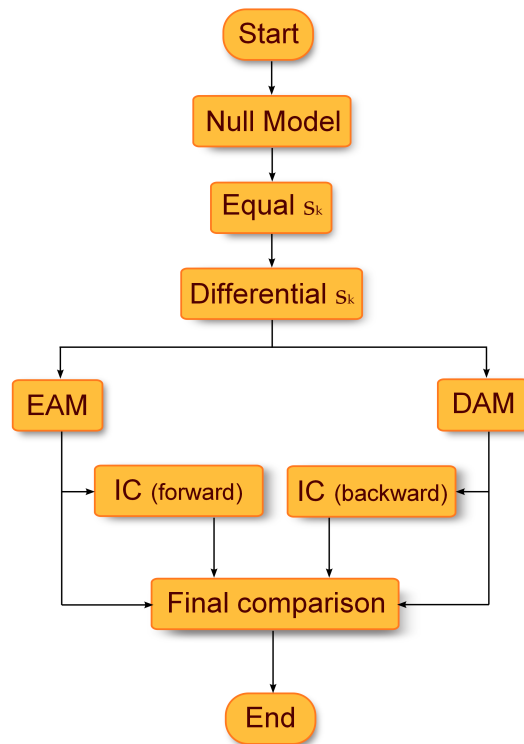


Figura 2.3: Flusso algoritmico della seconda versione di R-Average. Dopo la stima del modello nullo, i valori di scala vengono diversificati prima *tra* fattori e poi *entro* fattori. Il miglior modello selezionato passo dopo passo verrà utilizzato come punto di partenza per le procedure EAM, DAM e IC.

fine, partendo dal miglior modello selezionato, vengono stimati i modelli EAM, DAM e IC.

La nuova metodologia permette quindi di verificare l'esistenza di una differenza prima fra i valori scala (sia entro che tra fattori) e poi fra i pesi. Un confronto tra il modello con s differenziali e peso fisso, e il modello EAM che propone anch'esso s differenziali ma anche pesi differenziali tra fattori, permetterà d'identificare differenze fra il peso associato a fattori diversi.

I successivi passaggi procedurali sono analoghi a quanto esposto in precedenza, con la differenza che ora l'ottimizzazione non viene più effettuata sui pesi intesi come parametri w ma come parametri t .

2.9 Nuovo studio Monte Carlo

2.9.1 Metodo

Per testare la performance della nuova versione di R-Average, è stato realizzato un nuovo studio Monte Carlo. Il metodo utilizzato è stato del tutto analogo a quello precedente, per cui si rimanda al paragrafo 2.6.1 per i dettagli. Inoltre, sono stati utilizzati gli stessi dati sintetici generati in precedenza.

Le simulazioni sono state eseguite utilizzando il pacchetto *rAverage* versione 0.4-1 (appendice B), applicando sempre il metodo di minimizzazione L-BFGS-B, limitando le variazioni dei parametri s nell'intervallo $[0, 20]$ e le variazioni dei parametri t nell'intervallo $[-4, +4]$. Il valore di partenza fissato per i valori di scala era $s = 10$. La differenza minima per accettare l'uguaglianza fra due pesi è stata impostata pari a $\Delta_t = 0.1$, e la differenza minima fra i criteri dell'informazione necessaria per sancire l'uguaglianza fra due modelli è stata impostata pari a $\Delta BIC = \Delta AIC = 2$.

Le stime sono state realizzate utilizzando soltanto la versione completa della procedura Information Criteria, testando quindi tutte le possibili combinazioni di pesi. Diversamente dal precedente studio, però, per i gruppi di parametri 1 e 3 (tabella 2.1) la stima è stata effettuata anche eliminando dalla computazione le costanti di stato iniziale.

2.9.2 Risultati

Le tabelle A.9-A.12 (appendice A) mostrano i risultati ottenuti. A differenza del precedente studio, l'intestazione delle tabelle riporta il valore dei parametri t normalizzati.

Per quanto riguarda i valori di scala, le stime appaiono anche stavolta buone, e le distribuzioni dei parametri risultano simili a quelle che si osservavano con la precedente versione del metodo. Anche per i pesi, i risultati possono considerarsi analoghi, con una buona precisione nell'identificazione delle uguaglianze entro e tra fattori, ma una minore precisione nell'identificazione del valore puntuale del parametro rispetto ai valori di scala.

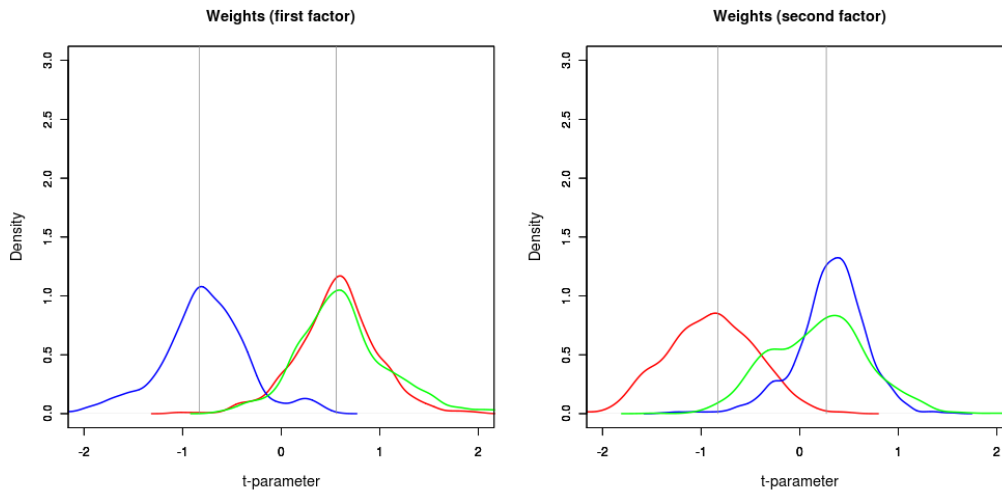


Figura 2.4: Distribuzioni dei parametri t per la simulazione realizzata sul primo gruppo di parametri per il caso 5 replicazioni e deviazione standard d'errore $\sigma_\epsilon = 1.0$. Per ogni fattore, le tre curve rappresentano le distribuzioni dei parametri di ogni livello (rosso: primo livello, blu: secondo livello, verde: terzo livello). Le linee verticali grigie rappresentano il valore reale del parametro.

A dispetto di quanto accadeva con la precedente versione del metodo per i w , ora anche la distribuzione dei t appare vicina alla simmetria. In figura 2.4 sono riportate le funzioni di densità per un caso esemplificativo (primo gruppo di parametri); si osserva che le distribuzioni dei t relativi ai livelli 1 e 3 per il primo fattore, e 2 e 3 per il secondo fattore, appaiono sovrapposte. Questi parametri sono uguali e difatti le loro distribuzioni risultano centrate intorno allo stesso valore.

Quando le condizioni di stato iniziale sono nulle ma comunque considerate nel modello, come nella precedente versione del metodo queste sono correttamente stimate intorno alle zero (tabella A.13). Quando la varianza d'errore sui dati è contenuta, il valore e^{t_0} è individuato come minore di 0.1 almeno nel 78% dei casi. Tuttavia, questa percentuale cala drasticamente all'aumento della varianza d'errore, e in misura maggiore rispetto a quanto accadeva con la precedente versione del metodo⁴.

⁴Verosimilmente, la differenza rispetto alla precedente versione del metodo può risiedere nei diversi limiti utilizzati per vincolare la routine di minimizzazione nel processo di stima di w_0 : $[1e - 5, 2]$ prima, $[e^{-4}, e^4]$ ora.

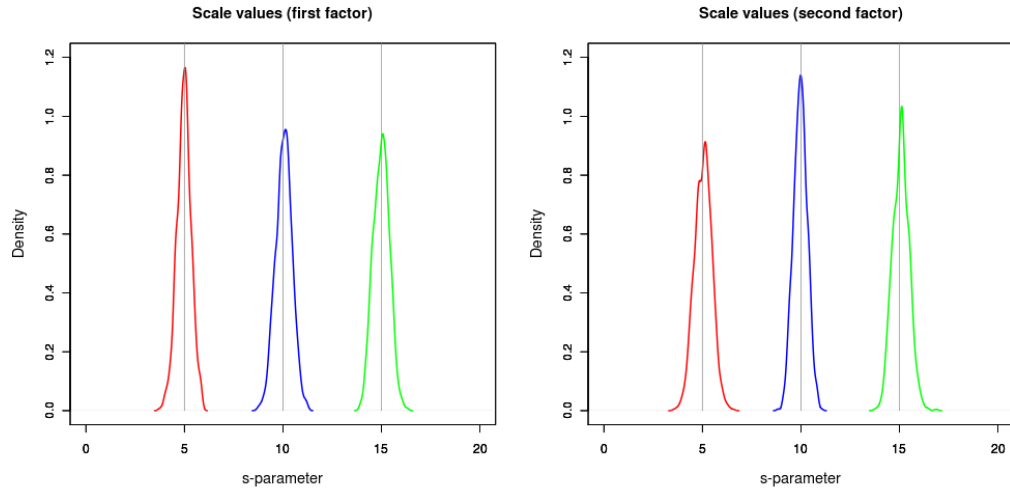


Figura 2.5: Distribuzioni dei parametri s per la simulazione realizzata sul primo gruppo di parametri eliminando dalla stima lo stato iniziale, assumendo 5 replicazioni, e $\sigma_\epsilon = 1.0$. Per ogni fattore, le tre curve rappresentano le distribuzioni dei parametri di ogni livello (rosso: primo livello, blu: secondo livello, verde: terzo livello). Le linee verticali grigie rappresentano il valore reale del parametro.

Quando le condizioni di stato iniziale vengono eliminate dal modello, le stime appaiono più precise (tabelle A.14-A.15). Consideriamo, a titolo esemplificativo, il caso in cui $\sigma_\epsilon = 1.0$ con 5 trial di stimolazione. Questa condizione è infatti quella più simile alle reali condizioni di applicazione del metodo R-Average, in quanto il numero di replicazioni è ridotto e la deviazione standard dell'errore non è né troppo bassa né troppo elevata⁵. Prendiamo in particolare il primo gruppo di parametri, che presenta condizioni iniziali nulle e pesi parzialmente diversi entro fattore per entrambi i fattori, con $t_A = \{0.56, -0.83, 0.56\}$ e $t_B = \{-0.83, 0.27, 0.27\}$.

In figura 2.5 sono riportate le distribuzioni degli s , stimati eliminando dal modello le costanti di stato iniziale (il cui valore è appunto nullo). Le curve appaiono centrate sul valore reale del parametro, con una conformazione perfettamente simmetrica e delle deviazioni standard molto ridotte ($\sigma < 0.4$ in tutti i casi). Le distribuzioni appaiono migliori di quelle che si osservano quando le condizioni iniziali vengono comunque considerate nella stima

⁵è stato verificato su dati sperimentali, come quelli presentati nel cap. 3, che in media la deviazione standard d'errore sui dati del singolo soggetto per una scala 0-20 è circa 1.0.

nonostante siano in realtà nulle⁶.

Le distribuzioni dei t [figura 2.6(a)] rimangono centrate in un punto prossimo al valore reale del parametro e vicine alla simmetria, anche se la forma non è eccellente come quella dei valori scala. Si nota comunque che aumentando il numero di replicazioni da 5 a 10 [figura 2.6(b)] le forme delle distribuzioni migliorano notevolmente: la varianza si riduce e la curva si avvicina maggiormente alla forma tipica della gaussiana.

2.9.3 Discussione

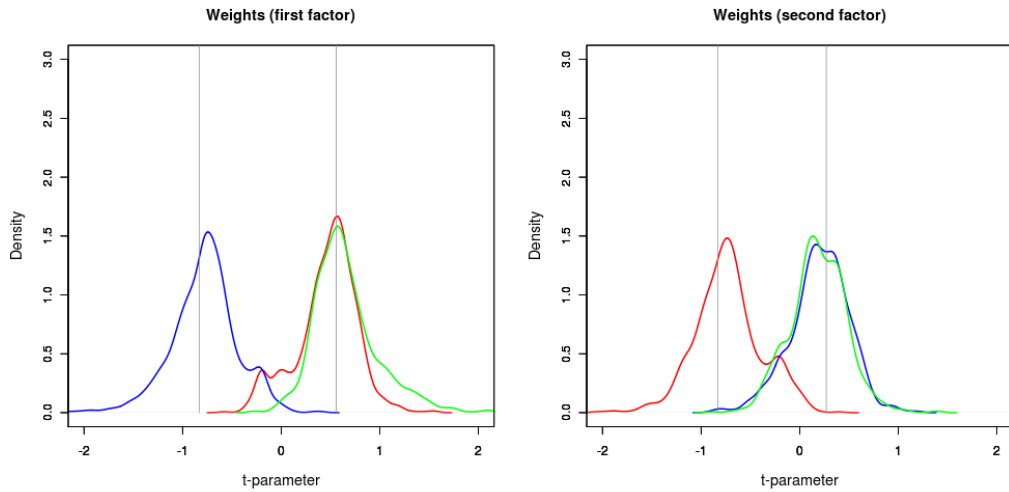
I risultati dello studio Monte Carlo realizzato con la seconda versione del metodo confermano quanto osservato in precedenza.

I valori di scala risultano ben stimati, con delle distribuzioni normali centrate sui valori reali dei parametri. Inoltre, se lo stato iniziale è nullo, queste distribuzioni migliorano notevolmente se nel processo di stima le costanti s_0 e w_0 vengono eliminate. Molte applicazioni sperimentali effettivamente non considerano lo stato iniziale, e R-Average fornisce la possibilità di non computarle.

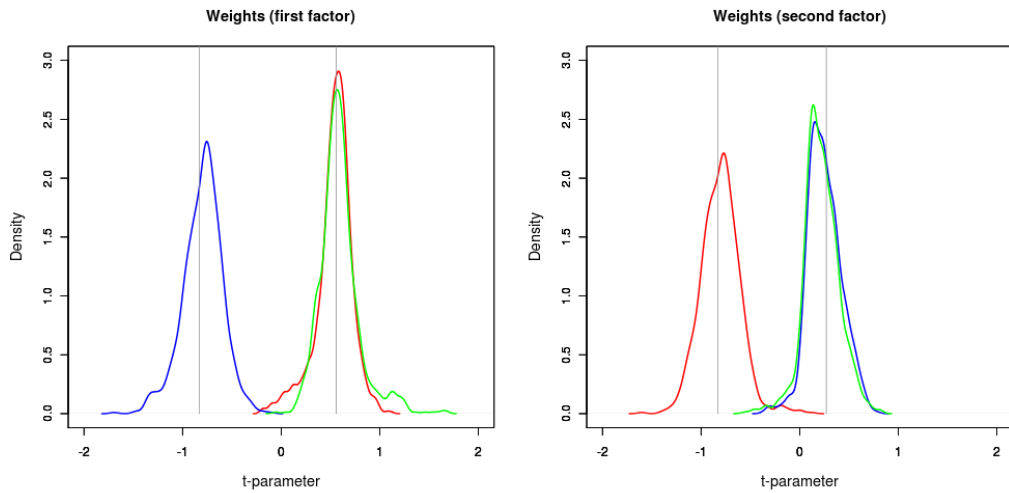
Con la nuova versione del metodo, anche per i pesi, concepiti come t e non più come w , si osservano delle distribuzioni vicine alla simmetria. Tuttavia, in futuro saranno necessari ulteriori miglioramenti del metodo di stima per eliminare definitivamente problemi di minimo locale che ancora si possono verificare. In figura 2.6(a) si osservano infatti distribuzioni non perfettamente simmetriche, con dei punti lontani dal valore reale del parametro in cui la densità tende a concentrarsi in maniera anomala. In ogni caso, aumentare il numero di replicazioni sul singolo soggetto aiuta la convergenza delle procedure, che con minore probabilità ricadranno in un minimo locale.

A questo punto, il metodo R-Average può considerarsi maturo e pronto per essere applicato per l'analisi dei dati in studi sperimentali.

⁶Si confrontino le curve rappresentate in figura 2.5 con quelle in figura 2.2 (come detto, se lo stato iniziale è presente, le distribuzioni degli s sono simili a quelle del precedente studio).



(a) Simulazione realizzata sui dati con 5 replicazioni.



(b) Simulazione realizzata sui dati con 10 replicazioni.

Figura 2.6: Distribuzioni dei parametri t per le simulazioni con $\sigma_\epsilon = 1.0$, con 5 replicazioni [2.6(a)] e 10 replicazioni [2.6(b)] realizzate sul primo gruppo di parametri eliminando dalla stima lo stato iniziale. Per ogni fattore, le tre curve rappresentano le distribuzioni dei parametri di ogni livello (rosso: primo livello, blu: secondo livello, verde: terzo livello). Le linee verticali grigie rappresentano il valore reale del parametro.

Capitolo 3

Il modello averaging nelle ricerche di marketing

3.1 Introduzione

Nell'acquistare un prodotto, implicitamente i consumatori esprimono una preferenza. Nel campo del marketing, la comprensione di come queste preferenze sono formulate e da quali componenti vengono modulate è un argomento particolarmente rilevante per il posizionamento di un prodotto sul mercato. Delineare il profilo di un prodotto in modo da renderlo quanto più attraente possibile risulta quindi una questione di vitale importanza.

Ogni prodotto presenta diverse caratteristiche che possono essere delineate in modi molto diversi, generando un impatto psicologico diverso sul consumatore. Per esempio, il colore della confezione potrebbe apparentemente sembrare un'inutile caratteristica perché slegato dall'effettiva qualità del prodotto, ma giocare in realtà un ruolo importante nell'attrarre la preferenza del consumatore (Bellizzi e Crowley, 1983).

Inoltre, prima d'inserire un nuovo prodotto sul mercato, predire l'impatto che ogni sua caratteristica avrà sulla percezione dei possibili acquirenti è una questione importante, perché permette di calibrarne adeguatamente le caratteristiche su determinate fasce di consumatori.

La Teoria dell'Integrazione delle Informazioni descrive la valutazione di

una caratteristica di un bene scindendone il valore dall'importanza che questa ha nel determinare il giudizio. Il modello averaging assume che caratteristiche diverse potrebbero incidere diversamente sulla percezione che ne hanno i consumatori, a prescindere dal valore che queste assumono. Un esempio è fornito dalla ricerca di Jo et al. (2003), che mostra come una famosa marca di televisori possa essere considerata poco attraente rispetto ad altre marche meno rinomate quando è specificata "Singapore" come località di produzione dell'oggetto. La valutazione di un prodotto, quindi, potrebbe non derivare dalla mera integrazione additiva dei valori attribuiti alle sue componenti. Le valutazioni delle singole componenti potrebbero influenzarsi vicendevolmente, dando origine a effetti d'interazione.

Per via delle sue peculiarità, il modello averaging presenta un grosso potenziale che può essere pienamente sfruttato nelle ricerche di marketing. Diversi studi hanno infatti mostrato che spesso, in questo campo, il processo d'integrazione può essere ben descritto da un modello averaging (per es.: Troutman e Shanteau, 1976; Johar, Jeedidi e Jacoby, 1997; Dougherty e Shanteau, 1999). Attraverso un sistema di ponderazione differenziale dei livelli dei fattori, il modello permette di tenere in considerazione il fatto che le valutazioni dei singoli attributi possono variare in funzione di come i vari livelli sono fra loro combinati.

Per esempio, un prezzo basso o un prezzo elevato potrebbero presentare un'importanza superiore rispetto a un prezzo di media entità. In generale, un livello di un fattore a cui è associato un valore di scala estremo (sia in senso positivo che negativo) potrebbe presentare un'importanza maggiore nel giudizio complessivo finale (Skowronski e Carlston, 1987; Jo et al., 2003). Inoltre, alcuni studi hanno mostrato che uno stesso livello di valutazione media può decrementare il giudizio complessivo se sommato a un livello dal valore elevato, ma incrementarlo se sommato a un livello di basso valore (Anderson, 1973). Questo fenomeno può avere notevoli ripercussioni sul comportamento del consumatore. Per esempio, nel *dilution effect* (Meyvis e Janiszewski, 2002), l'aggiunta di informazioni irrilevanti può indebolire l'impatto dell'informazione principale, "diluendo" la risposta.

Ancora, è stato dimostrato che le persone percepiscono un valore superio-

re in uno sconto operato su un prezzo che è al di sopra del loro prezzo target rispetto a uno sconto applicato su un prezzo che ne è al di sotto (Janiszewski e Cunha, 2004). Quindi, il valore dello sconto influenza diversamente la percezione a seconda dell'entità del prezzo di riferimento. Nella pratica, questo effetto differente corrisponde a un'interazione tra fattori, e un modello averaging con una ponderazione differenziale per i livelli di uno stesso fattore può spiegare il fenomeno. In presenza di effetti d'interazione, infatti, una rappresentazione a due parametri può risultare più appropriata per identificare il reale contributo di ogni attributo dello stimolo nella formulazione del giudizio finale.

Il presente studio è stato condotto in collaborazione con un pastificio veneto, l'azienda Sgambaro, al fine di identificare le caratteristiche di un suo prodotto di punta su cui far leva per incrementare la valutazione del bene da parte dei consumatori. Ponendo particolare attenzione al confezionamento del prodotto, lo studio ha dato l'opportunità di esplorare le possibilità d'uso del modello averaging per studiare le preferenze dei consumatori. Il prodotto "pasta" è stato scelto in quanto ampiamente diffuso e noto in tutto il territorio italiano da una vastissima fascia di consumatori, e il suo mercato potrebbe considerarsi esemplificativo di tanti altri casi di prodotti alimentari.

Alla base della formulazione del giudizio relativo ai prodotti proposti è stata assunta una funzione d'integrazione di tipo averaging. Grazie all'uso di questo modello si cercherà di andare incontro alla necessità di valutare gli effetti d'interazione, spesso completamente ignorati nelle ricerche di marketing che si avvalgono di metodologie di misurazione congiunta (Carmone e Green, 1981).

3.2 Metodo

Attraverso un disegno sperimentale, sono stati manipolati due fattori: il tipo di confezione (a tre livelli: scatola con finestra, scatola senza finestra, busta di plastica) e il prezzo (0.89 euro, 0.99 euro, 1.09 euro). La busta di plastica è il confezionamento standard proposto dall'azienda; lo studio ha quindi voluto analizzare altre due possibili confezioni (le scatole di cartone) com-



Figura 3.1: Confezioni del prodotto presentate nel corso dell’esperimento. Da sinistra: busta di plastica, scatola senza finestra, scatola con finestra.

parandole con quella standard. A differenza della scatola “senza finestra”, quella “con finestra” presenta una piccola finestra trasparente attraverso la quale l’acquirente può vedere il prodotto all’interno (figura 3.1).

3.2.1 Partecipanti

Alla ricerca hanno preso parte 48 persone (età media: 48.7 anni, dev. st.: 14.6), di cui 21 maschi e 27 femmine, che sceglievano di partecipare su base volontaria. Per rafforzare la validità ecologica della ricerca, i partecipanti sono stati reclutati all’entrata di un grosso centro commerciale di Padova.

3.2.2 Apparato e stimoli

Gli stimoli erano costituiti dalle immagini del prodotto (figura 3.1): confezioni di pasta Sgamaro da 500 grammi ognuna. Queste immagini presentavano un tipo di confezione alla volta e un’etichetta che ne riportava il prezzo. I nove *profili* risultanti dal disegno fattoriale 3×3 costituivano il gruppo di stimoli del disegno fattoriale completo.

I tre tipi di confezione venivano inoltre presentati singolarmente, senza essere combinati con alcun livello di prezzo. Similmente, i tre prezzi venivano

presentati anche senza essere combinati con alcuna immagine del prodotto. I sei profili risultanti costituivano il gruppo di stimoli dei sotto-disegni a una via.

3.2.3 Procedura

I partecipanti sedevano davanti al monitor di un computer, posto in un apposito spazio dedicato in un locale del supermercato. Essi leggevano le istruzioni e iniziavano, dopo una breve sessione di prova, a valutare i profili del prodotto uno alla volta. I profili erano presentati in ordine randomizzato per blocchi (prima i profili dei sotto-disegni, poi quelli del disegno fattoriale completo); inoltre, essi venivano presentati per sei volte a uno stesso partecipante.

A ogni partecipante veniva chiesto di esprimere la propensione all'acquisto del prodotto; nel caso del sotto-disegno "solo prezzo", la richiesta era di quantificare la propensione all'acquisto di un pacco di pasta con il prezzo proposto (quindi a prescindere da altri tipi d'informazione).

Le risposte venivano espresse su una scala da 0 a 20, dove 0 indicava che non si sarebbe mai acquistato il prodotto e 20 indicava un acquisto sicuro. Le scale a 21 punti sono spesso utilizzate negli esperimenti di misurazione funzionale (si veda ad es. Anderson, 1972); esse infatti permettono di ottenere misure accurate, attendibili e valide, adeguate per l'analisi statistica dei dati (Pearse, 2011).

La scala di risposta era costituita da una linea orizzontale con delle tacche verticali in corrispondenza dei valori. La scala veniva posta al di sotto dell'immagine del prodotto, e il partecipante poteva selezionare col pulsante del mouse il valore desiderato e proseguire con l'esperimento. Tra uno stimolo e l'altro intercorreva un intervallo di due secondi; l'esperimento durava circa otto minuti.

Al termine della sessione sperimentale, ogni partecipante riceveva in regalo una confezione da 500 grammi di pasta.

3.3 Risultati

Come primo passo, i dati sono stati analizzati per soggetto singolo utilizzando la libreria *rAverage* versione 0.3-5. Assumendo come nulle le costanti di stato iniziale, per ogni soggetto sono stati stimati i parametri del modello averaging, e l'adattamento dei modelli ai dati è stato valutato in termini di R^2 . Quattro soggetti sono stati eliminati perché presentavano un R^2 inferiore a 0.1.

Partendo dai parametri stimati su ogni singolo soggetto¹, i partecipanti sono stati suddivisi in sottogruppi utilizzando un algoritmo agglomerativo di cluster analysis (Struyf, Hubert e Rousseeuw, 1997). Sono stati identificati tre gruppi: cluster A ($n = 19$), cluster B ($n = 15$) e cluster C ($n = 10$).

Questi cluster risultano diversi nelle proporzioni tra maschi e femmine. L'analisi, eseguita attraverso l'adattamento di un modello lineare generalizzato, ha evidenziato che nel cluster A il rapporto fra il numero di femmine e il numero di maschi è superiore rispetto agli altri due cluster ($z = 3.11, p < 0.01$). Diversamente, i cluster B e C non differiscono nella proporzione fra i sessi ($z = 0.38, p = 0.70$). Non si registrano differenze significative nell'età dei soggetti appartenenti ai tre cluster ($F = 0.30, p = 0.74, \eta^2 = 0.01$).

L'ANOVA eseguita sui dati raggruppati per cluster evidenzia, per quanto riguarda i fattori principali, una significatività del fattore Prezzo in tutti e tre i cluster: sia A ($F_{(2,36)} = 9.52, p < 0.001, \eta^2 = 0.35$) che B ($F_{(2,28)} = 8.41, p < 0.01, \eta^2 = 0.38$) che C ($F_{(2,18)} = 12.10, p < 0.001, \eta^2 = 0.57$). Per quanto riguarda la Confezione, questa risulta significativa nei cluster A ($F_{(2,36)} = 3.29, p < 0.05, \eta^2 = 0.15$) e C ($F_{(2,18)} = 10.38, p < 0.01, \eta^2 = 0.54$), ma non nel cluster B ($F_{(2,28)} = 2.58, p = 0.09, \eta^2 = 0.16$); per quanto riguarda il cluster A, comunque, si osserva una dimensione dell'effetto molto ridotta.

L'interazione tra i due fattori risulta significativa nel cluster C ($F_{(4,36)} = 5.14, p < 0.01, \eta^2 = 0.36$) e non significativa sia nel cluster A ($F_{(4,72)} = 1.08, p = 0.37, \eta^2 = 0.06$) che nel cluster B ($F_{(4,56)} = 1.07, p = 0.38, \eta^2 = 0.07$). Questo indica che i soggetti raggruppati nei cluster A e B comples-

¹Al fine di uniformare la scala dei pesi tra soggetti diversi, i parametri w sono stati scalati dividendoli per la loro media geometrica (si veda par. 2.7).

Cluster	Attributo	s_1	s_2	s_3	w_1	w_2	w_3
A	Prezzo	12.0	9.7	7.7	1.0	1.0	1.0
	Confezione	11.3	7.8	10.1	0.7	0.7	0.7
B	Prezzo	15.9	13.9	11.8	1.0	1.0	1.0
	Confezione	11.5	13.6	16.5	1.0	1.0	1.0
C	Prezzo	7.7	6.0	4.2	1.0	1.0	2.0
	Confezione	13.0	0.7	5.2	1.0	1.0	1.0

Tabella 3.1: Valori di scala e pesi del modello averaging suddivisi per livello del fattore, stimati per ogni cluster. Codifica livelli: Prezzo: 1 = 0.89 €, 2 = 0.99 €, 3 = 1.09 €; Confezione: 1 = scatola con finestra, 2 = scatola senza finestra, 3 = busta di plastica.

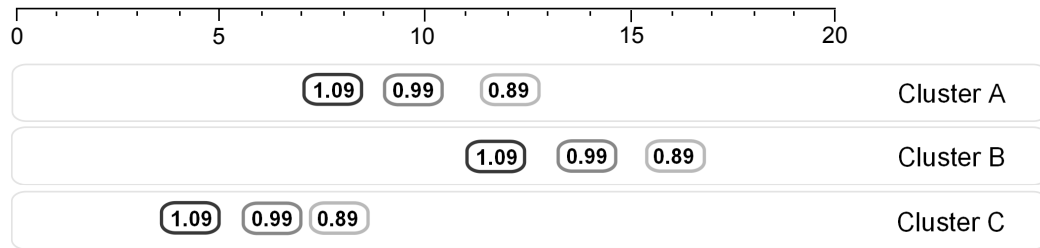
sivamente integrano i valori attribuiti ai livelli dei fattori in modo di fatto additivo, ponderando in maniera uguale tutti i livelli di uno stesso fattore. Diversamente, il cluster C dovrebbe mostrare un modello averaging a pesi diversi per almeno un fattore. La stima dei parametri del modello averaging sui dati raggruppati per cluster conferma questo risultato (tabella 3.1).

Per ciò che concerne i valori di scala, il cluster A esprime delle valutazioni che si collocano sul centro scala, il cluster B tende a fornire risposte sul versante positivo della scala mentre il cluster C fornisce delle valutazioni prevalentemente negative.

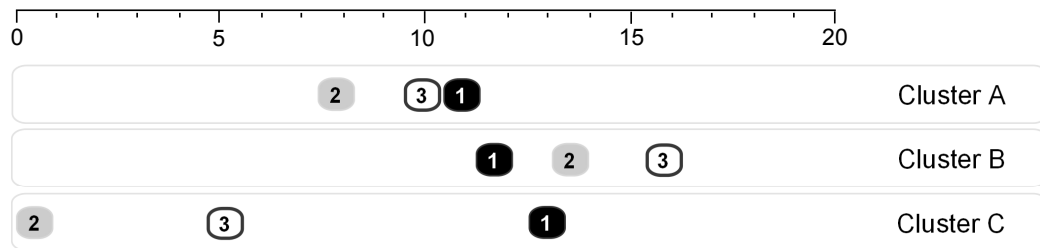
Per tutti i cluster, il prezzo più basso è quello preferito, mentre il prezzo più elevato è quello meno preferito [figura 3.2(a)].

Per quanto riguarda la confezione, i tre cluster evidenziano delle differenze nelle valutazioni, con delle preferenze diverse fra gruppi. Tuttavia, quello che si può notare è che la scatola con finestra, seppure non sia valutata come confezione migliore da tutti i cluster, raccoglie delle valutazioni che si posizionano sempre e comunque sul versante positivo della scala [figura 3.2(b)].

Per quanto riguarda i pesi, nel cluster C viene stimato un modello averaging a pesi diversi, dove il prezzo più elevato presenta un'importanza doppia rispetto agli altri livelli.



(a) Valori di scala per i livelli del fattore Prezzo.



(b) Valori di scala per i livelli del fattore Confezione.

Figura 3.2: Posizionamento dei parametri s sulla scala di risposta, per il prezzo (grafico in alto) e la confezione (grafico in basso), stimati per ogni cluster. Per la confezione, l’etichetta “1” identifica la scatola con finestra, “2” la scatola senza finestra e “3” la busta di plastica. I dati sono riportati in tabella 3.1.

3.4 Discussione

L’esperimento proposto è stato realizzato con lo scopo d’individuare quale tipologia di confezionamento di pasta risulta più attraente per il consumatore, valutando anche come i giudizi variano in funzione del prezzo attribuito al prodotto. Il modello averaging della IIT è stato utilizzato per suddividere i giudizi in un valore di scala e in un peso.

Le analisi dei dati hanno mostrato che la popolazione dei consumatori sembra non essere omogenea nelle modalità di valutazione del prodotto. Grazie alla metodologia R-Average, per ogni soggetto sono stati stimati i parametri del modello averaging, e sono state identificate tre diverse fasce di consumatori.

Sia per ciò che concerne i prezzi che i tipi di confezionamento del prodotto, un primo gruppo di consumatori esprime valutazioni “neutre”, che non si sbilanciano né verso polarità positive né negative. Un secondo gruppo esprime valutazioni che si collocano sul versante positivo della scala, e un terzo

gruppo valuta complessivamente negativamente gli attributi dei prodotti.

Per quanto riguarda il prezzo, in tutti i gruppi un costo elevato del prodotto viene considerato come un deterrente all'acquisto. Tuttavia, se in chi esprime valutazioni neutre e positive tutti i livelli di prezzo assumono sempre la stessa importanza, in chi esprime valutazioni negative il prezzo elevato assume un peso maggiore, per cui viene tenuto maggiormente in considerazione al momento della valutazione. Un peso elevato infatti accentua l'incidenza di un fattore rispetto all'altro. Quindi, quando un qualsiasi attributo verrà combinato con un prezzo elevato, il giudizio finale sarà determinato più dalla valutazione del prezzo che dell'altro attributo. In questi casi, associare un prodotto dal costo elevato con un altro attributo particolarmente piacevole, potrà aiutare a contrastare l'incidenza negativa del prezzo.

Fondamentalmente, quello che si deduce che è i consumatori che tendenzialmente non gradiscono il prodotto sono anche quelli che si fanno più influenzare da un costo elevato.

Per quanto riguarda il tipo di confezione, la ricerca ha messo in luce che la scatola con finestra è un modo "sicuro" di presentare il prodotto. Difatti, seppure in alcuni casi non risulti la tipologia di confezionamento considerata come migliore, per la distribuzione del prodotto nel mercato generale la scatola con finestra potrà assicurare comunque l'ottenimento di valutazioni positive del prodotto, raggiungendo un buon compromesso fra le preferenze di diverse tipologie di consumatori. Tuttavia, bisogna considerare che se il target di consumatori a cui punta la vendita non è la popolazione generale ma uno specifico sottogruppo, allora il prodotto dovrà essere definito sulla base delle specifiche valutazioni di questa fascia di consumatori.

Complessivamente, l'applicazione del modello averaging ha permesso di analizzare molto dettagliatamente i giudizi espressi dai partecipanti alla ricerca, permettendo di separare il valore di ogni attributo del prodotto dalla sua incidenza nella risposta complessiva. L'associazione del metodo R-Average con la cluster analysis ha permesso di segmentare il mercato e di delineare il profilo del prodotto sulla base di specifiche tipologie di consumatori target.

In ogni caso, ricerche di questo tipo richiedono numerosità campionarie ben maggiori rispetto a quella qui considerata. In futuro sarà quindi neces-

sario un ampliamento del campione, al fine di poter generalizzare maggiormente i risultati ottenuti. Sarà inoltre necessaria l'individuazione di ulteriori variabili, anche di tipo socio-economico, in grado di spiegare le differenze fra cluster di consumatori.

Capitolo 4

Verifica dell'ipotesi averaging nell'effetto di subadditività

4.1 Introduzione

La presente ricerca si colloca nell'ambito dello studio della formulazione dei giudizi in condizioni di rischio e incertezza. Secondo la classica Teoria dell'Utilità Attesa (Savage, 1954), una persona posta davanti a una presa di decisione che coinvolge alternative rischiose o incerte, valuterà le opzioni e opererà una scelta cercando di massimizzare il guadagno (Friedman e Savage, 1952).

L'utilità, cioè il valore complessivo di ogni opzione, è dato dal prodotto tra il valore implicito assegnato al bene che si può ottenere e il valore implicito assegnato alla probabilità di ottenimento del bene. Il giudizio relativo a una scommessa, descritta dal valore V di un bene e da una determinata probabilità P di ottenerlo, sarà quindi derivato attraverso un modello moltiplicativo (Anderson e Shanteau, 1970).

Uno dei principi fondanti la teoria dell'utilità attesa vuole che, in una situazione che coinvolge più opzioni, ognuna con il proprio valore e la propria probabilità di essere guadagnata, l'utilità complessiva derivi dalla somma di tutte le utilità singole. Quindi, il giudizio complessivo su una situazione che coinvolge K beni, ognuno con un valore V e una certa probabilità P di essere

ottenuto, deriverebbe da:

$$r = \sum_{i=1}^K P_i V_i \quad (4.1)$$

La formula (4.1) altro non è che un modello additivo, dove i valori di scala sono determinati dal prodotto tra il valore implicito del bene V_i e il valore implicito della probabilità P_i .

Tuttavia, già alla fine degli anni settanta quest'idea fu superata (Kahneman e Tversky, 1979). La regola additiva non trovò riscontro in diverse applicazioni sperimentali (Shanteau 1974; Lynch e Cohen, 1978). In particolare, Shanteau (1974) rilevò un effetto che definì *subadditività*.

L'autore propose a un gruppo di soggetti delle ipotetiche scommesse nelle quali il partecipante poteva vincere un certo tipo di beni con delle determinate probabilità, e un altro tipo di beni anch'essi con delle specifiche probabilità di vittoria. Ogni soggetto valutava sia le scommesse singole che quelle doppie. Se le scommesse singole presentavano delle situazioni in cui era coinvolto un unico tipo di bene, le scommesse doppie coinvolgevano entrambi i tipi di bene. Si osservò che la somma delle valutazioni delle scommesse singole non coincideva con la valutazione delle due scommesse considerate insieme. In particolare, si verificò che all'aumento del valore dei beni e delle probabilità di vittoria, i giudizi espressi per le scommesse doppie aumentavano progressivamente sempre meno rispetto a quanto non accadeva per la somma delle scommesse singole. In pratica, l'utilità globale sembrava essere meno della somma delle utilità parziali.

4.2 Possibili spiegazioni del fenomeno

Per spiegare l'effetto di subadditività fu ipotizzata l'esistenza di un modello d'integrazione di tipo averaging, con l'intervento di un processo di ponderazione differenziale fra le due scommesse (Lynch, 1979). Secondo quest'ipotesi, il modello corretto d'integrazione sarebbe descritto dalla formula:

$$r = \sum_{i=1}^K \frac{w_i P_i V_i}{\sum_{i=1}^K w_i} \quad (4.2)$$

dove r rappresenta l'utilità globale e le componenti $P_i V_i$ le singole utilità parziali. Per un caso a due fattori, il modello ipotizzato sarà:

$$r_{A_i B_j} = \frac{w_A P_{A_i} V_{A_i} + w_B P_{B_j} V_{B_j}}{w_A + w_B} \quad (4.3)$$

Lynch confrontò quest'ipotesi con altre possibili alternative. Difatti, seppure i suoi dati facevano presumere la presenza di un modello averaging (il test degli effetti opposti lo confermava) egli provò a confutare la possibilità che l'“anomalia” risiedesse non a livello della funzione d'integrazione \mathbb{I} ma a livello della funzione di risposta \mathbb{M} . Si prefigurarono quindi diverse alternative spiegazioni del fenomeno di subadditività.

La prima possibilità prevedeva che la formulazione del giudizio potesse essere descritta da un modello averaging a pesi completamente uguali (quindi puramente additivo) ma con una funzione di risposta non lineare. Lynch definì tale funzione di risposta come¹:

$$\mathbb{J}(r) = C_0 + C_1 r + C_2 r^2 + C_3 r^3 \quad (4.4)$$

La seconda possibilità prevedeva che la funzione d'integrazione \mathbb{I} fosse quella descritta dalla (4.2), con funzione di risposta lineare \mathbb{M} .

La terza possibilità prevedeva che il modello averaging descritto dalla (4.2) fosse valido, ma che la funzione di risposta potesse comunque risultare non lineare.

Per verificare quale delle ipotesi si potesse considerare valida, l'autore si avvale del criterio di *scale convergence* (Birnbaum e Veit, 1974). Il principio su cui si basa lo *scale convergence* è che i valori di scala sono indipendenti dalla funzione d'integrazione che il soggetto utilizza. Per questo motivo, al variare della funzione d'integrazione, i valori di scala dovrebbero rimanere

¹Nella presente trattazione, la notazione originale del lavoro di Lynch è stata parzialmente modificata per congruenza con i capitoli esposti in precedenza e con la terminologia adottata dallo stesso Anderson nei suoi articoli e manuali.

costanti. La metodologia proposta da Birnbaum e Veit prevede di “forzare” nei soggetti l’uso di specifiche regole d’integrazione, e poi di confrontare i valori di scala misurati sotto le diverse condizioni d’integrazione².

Si potrebbe ad esempio chiedere esplicitamente ai soggetti di utilizzare nella valutazione prima una regola additiva e poi moltiplicativa (o viceversa). I valori di scala misurati sotto la condizione della regola moltiplicativa dovrebbero risultare sovrapponibili a quelli misurati sotto la condizione della regola additiva, a meno di un riscaldamento³. Quindi, dopo un opportuno riscaldamento dei parametri s , se la funzione di risposta è lineare allora si osserverà una relazione lineare tra i valori di scala ottenuti attraverso le due regole d’integrazione. Diversamente, se la funzione di risposta non è lineare, allora lo scale convergence fallirà.

Lynch (1979) applicò questa metodologia, proponendo ai soggetti due compiti: uno d’integrazione tra due tipi di scommesse, ognuna decritta da un valore e una probabilità, e uno di tipo sottrattivo. Se il primo compito chiedeva di valutare complessivamente l’offerta costituita da entrambe le scommesse, il compito sottrattivo prevedeva un confronto tra le due scommesse: il giudizio espresso era costituito dalla sottrazione dei loro valori.

I risultati portarono l’autore a rifiutare il modello additivo, in quanto il riscaldamento polinomiale (4.4) non permetteva di ottenere dei valori di scala congruenti col modello sottrattivo (la convergenza fra i valori di scala falliva). Diversamente, il modello averaging fu considerato valido, ma non fu possibile escludere completamente l’esistenza di una non linearità nella funzione di risposta: assumendo una regola d’integrazione averaging, il riscaldamento operato attraverso la funzione \mathbb{J} portava a una convergenza migliore di quello operato dalla funzione monotona \mathbb{M} . Un approfondimento ulteriore mai pubblicato (Lynch, 1979, p. 412, nota 8) lo portò comunque a escludere la non linearità della funzione di risposta.

²Come visto nel cap. 1, nei modelli additivi e moltiplicativi i valori di scala possono essere stimati dalle medie marginali.

³È possibile passare dal modello moltiplicativo a quello additivo attraverso una trasformazione logaritmica dei valori di scala, infatti: $\ln(s_{A_i} s_{B_j}) = \ln(s_{A_i}) + \ln(s_{B_j})$.

4.3 Lo studio proposto

Gli studi di Shanteau e Lynch si collocano in un'epoca in cui la misurazione funzionale iniziava, passo dopo passo, a maturare e a diffondersi. All'epoca ancora non esisteva alcun tipo di metodologia per la stima dei parametri del modello averaging, né tanto meno era stata proposta un'adeguata procedura per la selezione di un adeguato modello in grado di spiegare i dati osservati.

L'applicazione del criterio di scale convergence rispose parzialmente a questa richiesta metodologica degli sperimentatori; seppur valido e teoricamente fondato, lo scale convergence non poteva e non può rimpiazzare appropriate metodologie di stima. Lynch arrivò per via indiretta alla soluzione del problema, attraverso la comparazione delle risposte fornite in due esperimenti che proponevano compiti diversi. Il suo lavoro era fortemente penalizzato dal non possedere un adeguato metodo per la divisione dei parametri di scala dai pesi, che permetterebbe una valutazione diretta della congruenza della formulazione (4.2) con i dati osservati.

Oggi, il metodo R-Average (cap. 2) è in grado di chiarire il problema della subadditività in maniera definitiva, grazie a una stima dei parametri del modello effettuata direttamente della decomposizione delle risposte osservate. Infatti, se la formula (4.2) rappresenta il modello appropriato di formulazione del giudizio, e se la risposta esplicita è una trasformazione lineare della risposta implicita, allora la somma delle risposte osservate per le scommesse singole dovrebbe essere in relazione diretta col valore ottenuto dalla media ponderata delle stesse quando valutate simultaneamente.

Nel presente studio, una serie di ipotetiche scommesse, presentate sia singolarmente che in coppia, verrà valutata da un gruppo di soggetti. Attraverso la metodologia R-Average si stimerà il modello che meglio si adatta ai dati osservati, con lo scopo di verificare l'ipotesi di un'integrazione di tipo averaging nelle scommesse doppie. Inoltre, in due distinte condizioni sperimentali, si cercherà di manipolare l'importanza associata a una scommessa piuttosto che all'altra; se il modello d'integrazione è davvero averaging, ci si aspetta che il valore dei pesi vari congruentemente con la manipolazione sperimentale.

4.4 Metodo

4.4.1 Partecipanti

Alla ricerca hanno preso parte 30 soggetti (26 femmine, 4 maschi, età media: 20.6 anni), studenti di psicologia dell'Università di Coimbra, Portogallo. I soggetti partecipavano su base volontaria, e in cambio ricevevano crediti formativi universitari.

4.4.2 Disegno sperimentale

I soggetti venivano calati in un ipotetico scenario in cui un operatore telefonico proponeva una serie di scommesse nelle quali un particolare bene poteva essere ottenuto, con una data probabilità. Le scommesse proponevano due tipi di beni: telefoni cellulari oppure servizi telefonici consistenti in un numero giornaliero di messaggi SMS gratuiti.

I telefoni cellulari offerti potevano essere di tre tipi: uno di scarso valore (Phone A), uno di medio valore (Phone B) e uno di valore elevato (Phone C). Le scommesse sul servizio telefonico proponevano un guadagno di 5, 10 oppure 15 messaggi SMS gratis al giorno. Ogni bene poteva essere guadagnato con una probabilità del 15%, del 50% oppure dell'85%. Dalla combinazione delle $3 \times 3 = 9$ scommesse relative al telefono cellulare e delle $3 \times 3 = 9$ scommesse relative al servizio telefonico, sono state ottenute le $9 \times 9 = 81$ scommesse doppie. Le scommesse venivano proposte in due possibili contesti.

Nel contesto *Phone trust*, ai partecipanti veniva spiegato che i telefoni cellulari sono forniti da uno dei più prestigiosi fabbricanti europei, insignito dell'ultimo premio "Qualità e ottimizzazione", che di recente ha investito grossi capitali per rinforzare il processo produttivo. Parallelamente, ai partecipanti veniva spiegato che il servizio di messaggistica è fornito da una compagnia telefonica che possiede esigue quote di mercato e che recentemente è stata abbandonata da molti utenti a causa di una debole copertura di rete e scarsa chiarezza nei contratti telefonici.

Nel contesto *Service trust*, la situazione era opposta rispetto alla precedente. Ai partecipanti veniva spiegato che i telefoni cellulari sono forniti da

un fabbricante che nell'anno passato è stato investito dalla crisi economica, e l'azienda madre starebbe valutando la possibilità di chiudere l'industria perché non soddisfatta della catena produttiva e dalla qualità dei prodotti finali. Parallelamente, il servizio di messaggistica è fornito da un'importante compagnia telefonica in espansione, che rapidamente sta guadagnando ampie quote di mercato grazie alle tariffe convenienti e all'ottima qualità del segnale.

Attraverso la manipolazione della qualità dei beni offerti, si è cercato di manipolare il grado di fiducia che il soggetto può riporre su tali offerte, modificando quindi l'appetibilità dei prodotti.

4.4.3 Ipotesi

Validità del modello averaging

L'ipotesi che muove la ricerca è che se la regola d'integrazione è averaging, allora l'effetto di subadditività sparirà, e si osserverà una relazione lineare tra la somma dei valori attribuiti alle scommesse singole e la media pesata dei valori di scala stimati. Non è quindi la somma dei valori delle scommesse singole a predire i valori delle risposte doppie, ma la loro media ponderata. Se il modello è veramente averaging, si osserverà una scomparsa dell'effetto di subadditività quando si cercano di predire le risposte osservate per le scommesse doppie attraverso i parametri stimati per il modello averaging.

Inoltre, se il modello è davvero averaging, la manipolazione della fiducia nei due contesti dovrebbe avere un effetto sull'importanza attribuita a ognuna delle due scommesse, che si rifletterà sui parametri di peso. Se nella condizione Phone trust si dovrebbe osservare un peso maggiore delle scommesse relative al telefono cellulare rispetto al servizio di messaggistica, nella condizione Service trust si dovrebbe osservare l'effetto opposto, con una ponderazione maggiore per le scommesse relative al servizio SMS rispetto a quelle relative al telefono. La manipolazione della fiducia si rifletterà anche sul grafico delle medie marginali, in quanto si dovrebbe osservare una variazione reciproca del loro andamento fra i due contesti. Al crescere dell'effetto

del fattore posto sull'asse delle ordinate, si dovrebbe osservare una riduzione dell'inclinazione del fattore posto sull'ascissa, e viceversa.

Ancora, se il modello è averaging, anche se l'analisi della varianza non mostrerà un'interazione significativa tra i due tipi di scommessa, l'aggiunta del livello in cui uno dei due fattori è valutato indipendentemente dall'altro (quindi singolarmente) renderà manifeste le deviazioni dal parallelismo delle medie marginali⁴.

Linearità della funzione di risposta

Quando il giudizio viene espresso su un fattore singolo, non combinato con alcun altro tipo d'informazione, se si assumono le costanti di stato iniziale nulle, allora le risposte osservate R dovrebbero coincidere con i valori di scala s (par. 1.7.1).

Avendo quindi a disposizione i valori di scala misurati direttamente sulle scommesse singole, sarà possibile compararli con quelli stimati per il disegno fattoriale completo. Se la funzione di risposta è lineare, allora si osserverà una relazione lineare tra i valori di scala rilevati per le scommesse singole e quelli stimati per le scommesse doppie.

4.4.4 Apparato e stimoli

Gli stimoli erano presentati sul monitor di un computer sotto la forma di cartellini (figura 4.1). Il cartellino posizionato sulla sinistra descriveva la scommessa per il servizio SMS e quello posizionato sulla destra per il telefono cellulare. Ogni cartellino riportava sia il premio che la probabilità di ottenerlo.

La scala di risposta era posta sotto i cartellini, ed era costituita da un segmento continuo che andava da un minimo a un massimo di convenienza percepita per l'offerta. Nella sessione di prova, essa presentava alle estremità due ancore che rappresentavano rispettivamente il valore minimo e il valore massimo.

⁴Si tratta del test degli effetti opposti, si veda il par. 1.6.2.

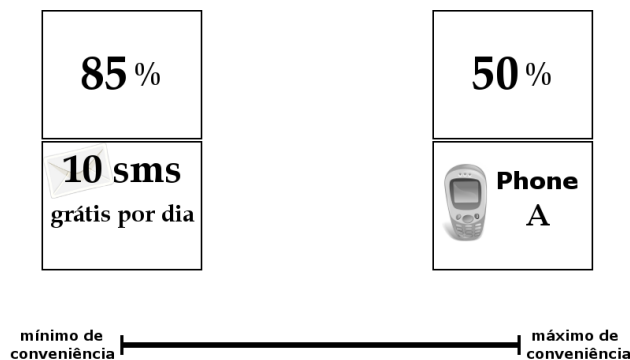


Figura 4.1: Esempio di stimolo presentato nel corso dell’esperimento per le scommesse doppie. A ognuno dei due possibili premi, telefono cellulare (posizionato a sinistra) e servizio SMS (posizionato a destra), era associata una probabilità di vittoria. I soggetti venivano preventivamente istruiti riguardo le caratteristiche dei telefoni cellulari che potevano vincere e addestrati all’uso della scala di risposta. Per le scommesse singole, solo uno dei due cartellini veniva presentato al centro dello schermo.

Le ancore erano costituite da dei premi di valore e probabilità leggermente inferiori a quelli degli stimoli da valutare nel caso del limite inferiore, e da un valore e una probabilità leggermente superiori rispetto agli stimoli da valutare nel caso del limite superiore. Le due ancore quindi presentavano delle offerte certamente meno convenienti (nel caso del minimo) e certamente più convenienti (nel caso del massimo) rispetto a tutti gli stimoli proposti.

4.4.5 Procedura

I partecipanti venivano fatti accomodare davanti al monitor del computer dove sarebbero stati presentati gli stimoli. Sul computer venivano presentate inizialmente delle istruzioni che descrivevano la prova. Uno sperimentatore introduceva i partecipanti all’esperimento, spiegando verbalmente i dettagli del compito e le caratteristiche delle offerte. terminate le istruzioni, venivano presentati dieci stimoli di prova attraverso i quali il partecipante poteva familiarizzare con il compito.

I soggetti venivano addestrati al corretto uso della scala di risposta. Infatti, se nella fase di training agli estremi della scala venivano presentate le ancore, nella sessione sperimentale queste non venivano più presentate (figura 4.1) per non appesantire il carico cognitivo.

Se per le scommesse singole veniva chiesto di valutare la convenienza della singola offerta, per le scommesse doppie l'offerta prevedeva la possibilità di poter vincere un certo servizio telefonico e contemporaneamente anche un certo telefono cellulare. In questi casi, quindi, la richiesta era di giudicare l'offerta complessivamente, considerando che entrambi i premi potevano essere vinti.

L'esperimento durava circa 20 minuti. Tutti i soggetti valutavano tutte le scommesse singole e doppie, in entrambi i contesti di fiducia (Phone trust/Service trust). L'ordine di presentazione dei due contesti era contro-bilanciato fra i partecipanti: metà dei soggetti affrontava prima il contesto Phone trust e poi Service trust, l'altra metà affrontava prima il contesto Service trust e poi Phone trust.

Gli stimoli erano presentati in ordine randomizzato all'interno di ogni contesto di valutazione. Per ogni contesto, quando la metà degli stimoli era stata presentata, compariva una schermata che ricordava al partecipante le ancore su cui calibrare la risposta.

Le analisi statistiche e grafiche sono state realizzate utilizzando il linguaggio R (R Development Core Team, 2011).

4.5 Risultati

4.5.1 Manipolazione della fiducia

Una prima analisi è stata condotta per verificare l'assenza di un effetto dell'ordine di presentazione dei contesti fra i partecipanti. Una ANOVA a due vie Ordine \times Contesto conferma che non c'è differenza fra gli ordini di presentazione ($F_{(1,28)} = 3.71, p = 0.06, \eta^2 = 0.11$) e che l'ordine di presentazione non influenza le risposte ai due contesti ($F_{(1,28)} = 0.09, p = 0.76, \eta^2 < 0.01$).

Per valutare se la manipolazione del grado di fiducia da attribuire alle scommesse abbia avuto effetto, è stata condotta una ANOVA sulle risposte fornite dai soggetti alle scommesse singole e due ANOVA sulle risposte fornite dai soggetti alle scommesse doppie (una per il contesto Phone trust e una per il contesto Service trust).

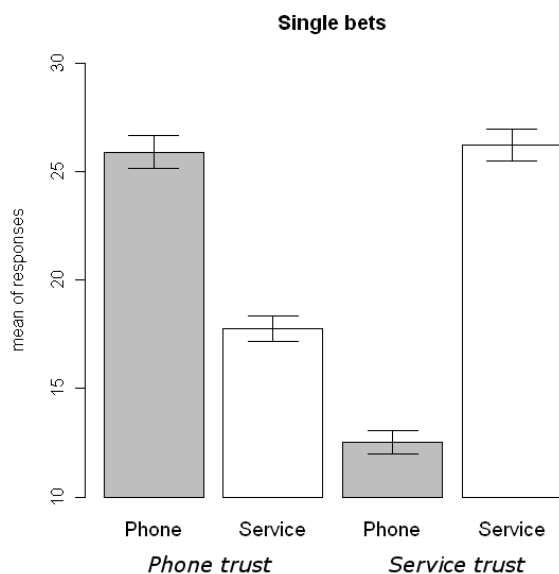


Figura 4.2: Medie marginali delle valutazioni espresse per le scommesse singole nei due contesti, per il telefono cellulare (in grigio) e il servizio SMS (in bianco). Le barre rappresentano gli errori standard.

Per ciò che concerne le scommesse singole, l'ANOVA è stata condotta utilizzando come fattori il tipo di scommessa e il contesto. I risultati evidenziano un'interazione significativa ($F_{(1,29)} = 69.11, p < 0.001, \eta^2 = 0.33$), che indica che la differenza fra le valutazioni delle due scommesse varia in funzione del contesto. Nella condizione in cui viene stimolata la fiducia sulle scommesse relative al telefono, si osserva un incremento di queste rispetto a quelle relative al servizio SMS. Quando invece viene stimolata la fiducia sul servizio SMS, la relazione si inverte e le preferenze sono maggiori per queste scommesse rispetto a quelle relative al telefono cellulare (figura 4.2).

Per le risposte doppie, l'ANOVA è stata condotta utilizzando come fattori i due tipi di scommessa, a nove livelli ognuno. Per il contesto Phone trust, i risultati mostrano che la manipolazione della fiducia, a favore del telefono cellulare, ha annullato l'effetto dell'offerta sul servizio SMS. L'ANOVA infatti evidenzia una significatività solo delle scommesse relative al telefono cellulare ($F_{(8,232)} = 114.14, p < 0.001, \eta^2 = 0.80$) e non di quelle relative al servizio SMS ($F_{(8,232)} = 1.95, p = 0.053, \eta^2 = 0.06$). L'intera-

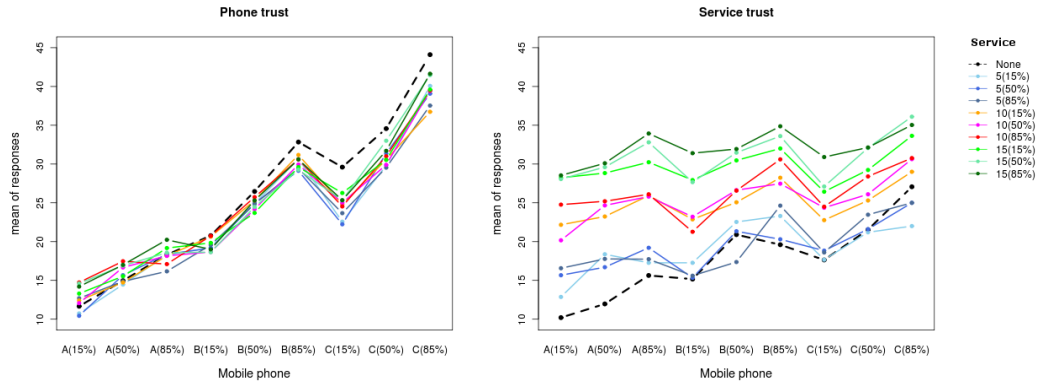


Figura 4.3: Medie marginali delle valutazioni espresse per le scommesse doppie nei due contesti. Per il contesto Phone trust, si osserva una sovrapposizione delle linee che congiungono le medie, a indicare la non significatività delle scommesse sul servizio SMS. Per il contesto Service trust si osserva un decremento nell’inclinazione delle linee ma un netto distacco fra esse, a indicare il decremento dell’effetto delle scommesse sul telefono cellulare e un incremento dell’effetto delle scommesse sul servizio SMS. Le linee nere tratteggiate congiungono le medie delle scommesse singole relative al solo telefono cellulare.

zione fra le due scommesse risulta non significativa ($F_{(64,1856)} = 1.09, p = 0.30, \eta^2 = 0.04$). Per il contesto Service trust, si osserva una significatività di entrambi i tipi di scommesse, sia quelle relative al telefono cellulare ($F_{(8,232)} = 15.19, p < 0.001, \eta^2 = 0.34$) che quelle relative al servizio SMS ($F_{(8,232)} = 40.37, p < 0.001, \eta^2 = 0.58$). L’interazione fra i due tipi di scommessa risulta significativa ($F_{(64,1856)} = 1.33, p < 0.05, \eta^2 = 0.04$); tuttavia, la dimensione dell’effetto risulta estremamente bassa, indicando uno scostamento dal parallelismo di entità trascurabile.

4.5.2 Validità del modello averaging

Per verificare la sussistenza di un modello averaging, è stata ricalcolata l’ANOVA sulle scommesse doppie aggiungendo al fattore “servizio SMS” un livello costituito dalla non presentazione dello stesso. Le scommesse singole, che costituiscono dei sotto-disegni in cui uno dei due tipi d’informazione è omessa, diventano quindi un ulteriore livello. Se il modello non è additivo ma averaging, ci si aspetta che l’aggiunta del fattore presentato singolarmente comporti una significatività dell’effetto interazione (par. 1.6.2). Infatti, la

formulazione delle risposte attraverso una regola averaging comporta che l'unione del valore attribuito al singolo fattore con ogni livello dell'altro fattore non porterà a un incremento costante della risposta.

L'ANOVA evidenzia ora un effetto d'interazione significativo sia per la condizione Phone trust ($F_{(72,2088)} = 1.91, p < 0.001, \eta^2 = 0.06$) che per la condizione Service trust ($F_{(72,2088)} = 1.98, p < 0.001, \eta^2 = 0.06$). Sembra quindi plausibile che il modello d'integrazione tra le scommesse doppie sia averaging. Tuttavia, la ridotta dimensione degli effetti d'interazione non può ancora fornire una conferma definitiva dell'ipotesi.

Per ogni soggetto sono stati stimati i parametri del modello averaging utilizzando la libreria *rAverage* versione 0.4-1, assumendo alla base della formulazione del giudizio un modello a pesi uguali.

Nella condizione Phone trust, per ventisei soggetti, il modello EAM a pesi uguali entro fattore ma diversi tra fattori, è stato selezionato come migliore rispetto agli altri. Il peso più elevato è, come atteso, assunto dalle scommesse relative al telefono cellulare. Per i restanti quattro soggetti, il modello migliore stimato prevede pesi uguali sia entro che tra fattori. In questi soggetti, quindi, la manipolazione della fiducia sembra non aver funzionato (le due scommesse assumono la medesima importanza nel giudizio finale).

Per la condizione Service trust, in quindici soggetti il modello a pesi uguali entro fattore ma diversi tra fattori è preferito rispetto agli altri. Tuttavia, di questi quindici, se in undici si osserva un peso superiore delle scommesse relative al servizio SMS, per quattro soggetti il peso maggiore è assunto sempre dalle scommesse relative al telefono cellulare. Dei restanti quindici soggetti, dieci attribuiscono uguale importanza alle due scommesse, e per cinque soggetti il modello nullo risulta migliore di tutti gli altri modelli stimati⁵. Fra i dieci soggetti per i quali viene stimato un modello a pesi completamente uguali, compaiono gli stessi quattro soggetti per i quali tale modello era stato stimato anche per la condizione Phone trust.

⁵Di questi cinque soggetti, ben quattro hanno affrontato prima il contesto Phone trust e poi Service trust. Verosimilmente, essendo la prova lunga e faticosa, arrivati al secondo contesto questi soggetti fornivano delle risposte poco coerenti per via della stanchezza, invalidando la prova.

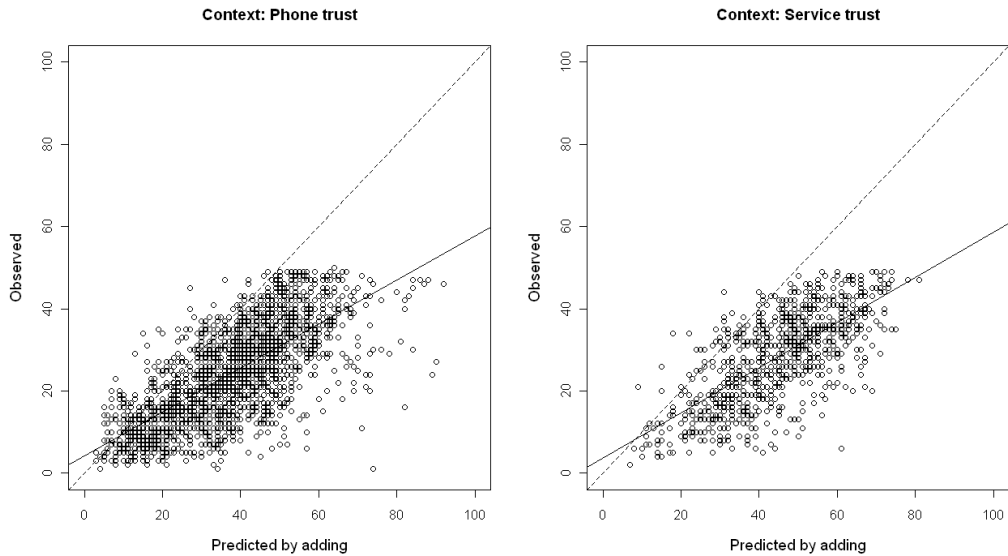


Figura 4.4: Valori previsti dal modello additivo (ottenuti dalla somma delle scommesse singole) contro i valori registrati per le scommesse doppie. L'effetto di subaddittività si manifesta col decremento dell'inclinazione dell'ideale retta di regressione (che ha uno slope pari a $\beta = 1$). Nei grafici sono considerati solo i soggetti per i quali la manipolazione della fiducia ha avuto effetto ($n = 26$ per Phone trust e $n = 11$ per Service trust).

Considerando per le due condizioni solo i dati dei soggetti dove la manipolazione della fiducia ha avuto effetto ($n = 26$ per Phone trust e $n = 11$ per Service trust), è stato adattato un modello di regressione lineare per ognuna delle due condizioni sperimentali, assumendo come variabile dipendente le risposte osservate per le scommesse doppie e come variabile indipendente la somma delle risposte osservate per le scommesse singole.

Se il modello additivo fosse la corretta regola d'integrazione, e la funzione di risposta fosse lineare, dovremmo aspettarci la stima di un modello in cui l'intercetta risulta pari a zero e il coefficiente angolare pari a uno. Tuttavia, i risultati evidenziano delle intercette significativamente diverse da zero e dei coefficienti angolari diversi da uno in entrambi i contesti (figura 4.4), sia per Phone trust ($\beta_0 = 4.11, t_{(2104)} = 8.70, p < 0.001; \beta_1 = 0.54, t_{(2104)} = 46.07, p < 0.001; R^2 = 0.50$) che per Service trust ($\beta_0 = 3.56, t_{(899)} = 4.00, p < 0.001; \beta_1 = 0.55, t_{(899)} = 28.94, p < 0.001; R^2 = 0.48$).

Se invece si assumono come variabili indipendenti le risposte previste

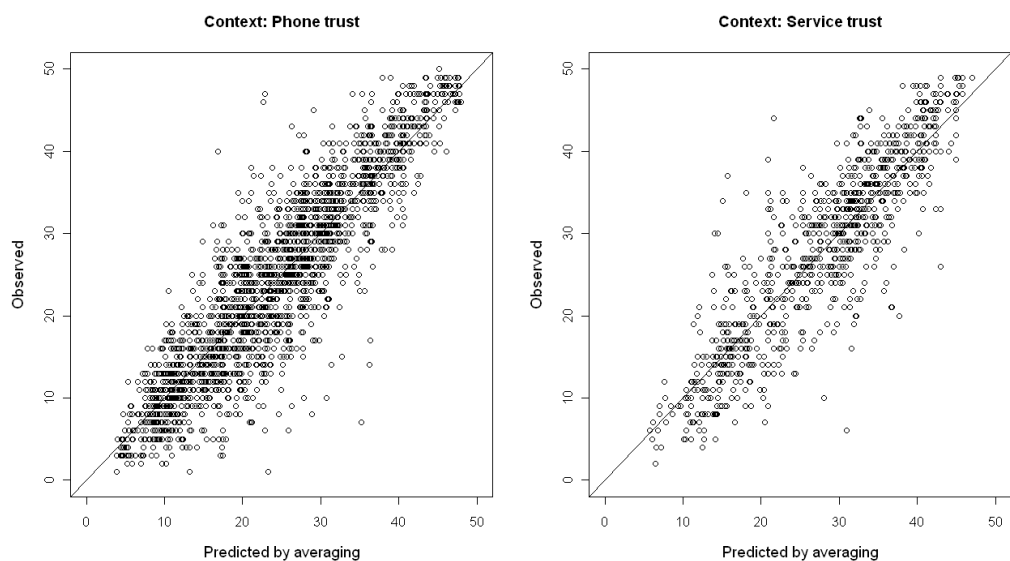


Figura 4.5: Valori previsti dal modello averaging (stimati tramite R-Average) contro i valori registrati per le scommesse doppie. Nei grafici sono considerati solo i soggetti per i quali la manipolazione della fiducia ha avuto effetto ($n = 26$ per Phone trust e $n = 11$ per Service trust).

attraverso il modello averaging, si osservano dei valori di intercetta non diversi da zero e dei coefficienti angolari molto vicini a uno (figura 4.5), sia per il contesto Phone trust ($\beta_0 = -0.26, t_{(2104)} = -0.93, p = 0.35; \beta_1 = 1.01, t_{(2104)} = 93.26, p < 0.001; R^2 = 0.80$) che per il contesto Service trust ($\beta_0 = -0.46, t_{(899)} = -0.84, p = 0.40; \beta_1 = 1.02, t_{(899)} = 55.28, p < 0.001; R^2 = 0.77$).

Per verificare che effettivamente i parametri di questi due modelli non differissero da 0 per l'intercetta e da 1 per il coefficiente angolare, il loro adattamento ai dati è stato confrontato con l'adattamento dei modelli dove i due parametri sono fissati rispettivamente pari a $\beta_0 = 0$ e $\beta_1 = 1$, assumendo quindi una relazione diretta tra variabile dipendente e indipendente. Questi due modelli a parametri fissi presentano gli stessi indici R^2 dei modelli stimati, per cui i due modelli stimati si possono considerare totalmente sovrapponibili a quelli a parametri fissi.

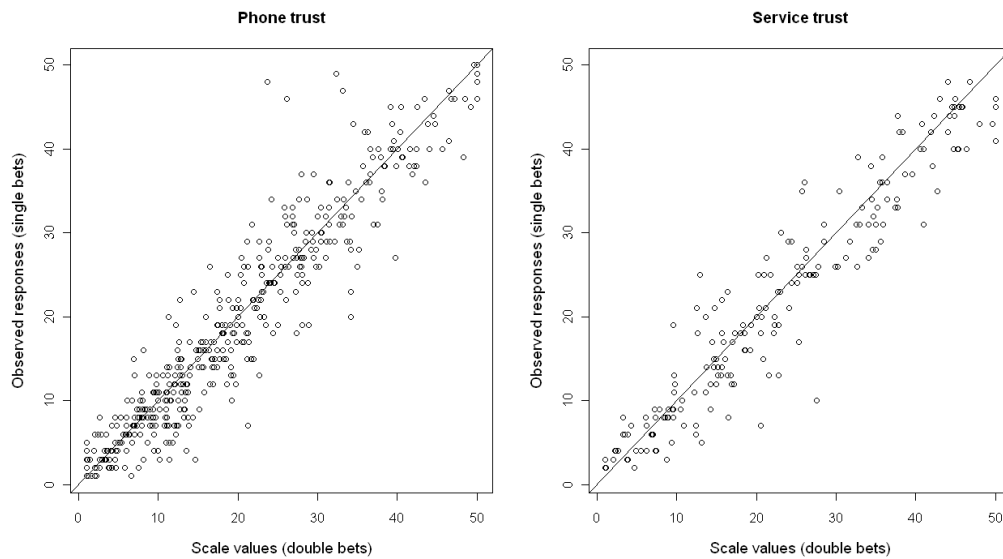


Figura 4.6: Valori di scala stimati tramite R-Average per le scommesse doppie contro i valori di scala ottenuti direttamente dalle risposte osservate per e scommesse singole.

4.5.3 Linearità della funzione di risposta

Al fine di valutare la sussistenza di una relazione lineare tra i valori di scala stimati per il compito comprendente le scommesse doppie e le risposte fornite per le scommesse singole, è stato adattato un modello di regressione lineare per ogni contesto. Anche in questo caso, sono stati considerati solo i soggetti sui quali la manipolazione della fiducia ha avuto effetto. I risultati mostrano che i valori di intercetta per entrambe le condizioni non risultano diversi da zero, e il valore dei coefficienti angolari si avvicina a uno, sia per il contesto Phone trust ($\beta_0 = 0.25, t_{(466)} = 0.71, p = 0.47; \beta_1 = 0.97, t_{(466)} = 63.31, p < 0.001; R^2 = 0.90$) che per il contesto Service trust ($\beta_0 = 0.27, t_{(196)} = 0.60, p = 0.55; \beta_1 = 0.98, t_{(196)} = 48.91, p < 0.001; R^2 = 0.92$).

L'assunzione dei valori $\beta_0 = 0$ e $\beta_1 = 1$ non modifica il valore dell'indice R^2 , per entrambi i contesti, per cui i due parametri si possono considerare sostanzialmente uguali a 0 e 1. Quindi, è confermata la relazione lineare tra i valori di scala delle scommesse doppie e singole, per cui la funzione di risposta può considerarsi a tutti gli effetti lineare.

4.6 Discussione

I dati presentati in questo capitolo supportano l'ipotesi averaging per l'effetto di subadditività. Quando le persone sono chiamate a valutare un'offerta costituita da due premi, ognuno dei quali può essere ottenuto con una data probabilità, le persone non sembrano effettuare una semplice somma tra le valutazioni dei due beni considerati singolarmente. Piuttosto, sembra intervenire un processo di ponderazione delle offerte. Difatti, quando viene diminuita l'attrattività di un bene, è il valore dell'altro a determinare il giudizio in maniera quasi esclusiva. Grazie all'applicazione del modello averaging, si è visto che effettivamente il peso associato al bene su cui è data fiducia aumenta, a scapito dell'altro.

Questo accade particolarmente quando è il bene “telefono cellulare” a essere valorizzato: in tale condizione, in ben ventisei soggetti su trenta si osserva questo fenomeno. Il risultato è invece meno evidente quando viene valorizzato il bene “servizio SMS”: in questo caso solo in undici partecipanti si osserva una ponderazione superiore per il peso associato a questo tipo di premi. Una spiegazione di questo differente trattamento che i soggetti riservano ai due beni può risiedere nel fatto che il telefono cellulare, essendo un bene tangibile e immediatamente riconoscibile, risulta più saliente⁶ rispetto al servizio telefonico. Quindi, accade che attribuendo fiducia al servizio SMS effettivamente se ne aumenti il peso, ma il telefono cellulare è così saliente da restare in tanti casi il fattore preponderante nel determinare il giudizio.

Lo studio ha anche messo in luce come la funzione di risposta sia lineare. Difatti, i valori di scala stimati per il modello a due fattori sono in relazione lineare diretta con le utilità delle scommesse singole. Quindi, diversamente da quanto aveva ipotizzato Lynch (1979), per spiegare il fenomeno della subadditività non risulta necessario chiamare in causa alcuna violazione dalla linearità della funzione di risposta.

⁶Il termine “saliente” è usato in questo caso secondo la definizione di *salience* fornita da Myers e Alpert (1977); si veda par. 1.4.

Capitolo 5

Il modello averaging nella valutazione delle credenze di fiducia

5.1 Introduzione

La fiducia è una componente fondamentale nel determinare le interazioni fra persone e organizzazioni, ma spesso è stata descritta e valutata in maniera approssimativa.

Rotter (1967) la definisce come l'aspettativa di una persona o di un gruppo di persone (*trustor*) che parole, promesse, affermazioni verbali o scritte di un altro gruppo di persone (*trustee*) si possano realizzare concretamente. Fondamentalmente, la fiducia si riferisce a come le persone esperiscono la tensione causata dalla scelta di dipendere dal comportamento degli altri invece di dirigere direttamente il proprio comportamento, per assicurarsi di non rimanere danneggiati. Partendo da queste assunzioni, Rotter (1967) sviluppò una scala per la misurazione della fiducia interpersonale.

Bromiley e Cummings (1995) combinarono tre componenti (cognitiva, affettiva e comportamentale) che precedono il comportamento umano e descrissero la fiducia organizzativa come la credenza di un individuo (o la una credenza comune di un gruppo) secondo la quale un altro individuo (o grup-

po) si sforza di agire in accordo con qualsiasi promessa manifesta o implicita, è sincero nelle discussioni davanti a tale promessa e non si appropria di benefici a lui non dovuti anche se si dovesse presentare l'occasione. Su queste basi, gli autori svilupparono un inventario per misurare la fiducia organizzativa (Cummings e Bromiley, 1996; Vidotto, Vicentini, Argentero, e Bromiley, 2008).

McKnight, Cummings, e Chervany (1998) e McKnight e Chervany (2000) proposero diverse tipologie di concetti che si possono riferire alla fiducia:

- credenze di fiducia, considerate come una profonda persuasione che il trustee abbia attributi favorevoli nell'indurre l'intenzione a fidarsi di lui;
- intenzioni di fiducia, considerate come una solida volontà a dipendere dal trustee, che induce comportamenti di fiducia;
- comportamenti di fiducia, considerati come azioni in grado di dimostrare che il trustor conta sul trustee senza operare alcun tipo di controllo;
- disposizione a fidarsi, considerata come una differenza individuale in una generale tendenza a riporre fiducia negli altri;
- fiducia basata sull'istituzione, considerata come un contesto favorevole in grado di supportare la fiducia.

Sia la disposizione a fidarsi che la fiducia basata sull'istituzione promuovono le credenze di fiducia, le intenzioni e i comportamenti.

Le credenze di fiducia sono fattori rilevanti nel fare in modo che un individuo consideri un'altra persona come degna di fiducia. Esse modulano il grado di sicurezza che un individuo ha nel ritenere che la persona su cui ripone la sua fiducia ne sia effettivamente degna. Secondo McKnight et al. (1998), le più citate credenze di fiducia sono:

- Benevolenza. Questa credenza viene definita attraverso l'interesse che una persona manifesta nei confronti di un'altra e la motivazione che la spinge ad agire nel suo interesse. Il trustee è concepito come una persona che ha buone intenzioni rispetto agli interessi e agli obiettivi del trustor e che non agisce opportunisticamente.

- **Competenza.** Questa credenza viene definita come la capacità di fare per un'altra persona quello di cui questa ha bisogno. Il trustee è quindi concepito come una persona che possiede le capacità e le abilità di cui il trustor ha bisogno.
- **Onestà.** Questa credenza viene definita come la capacità di prendere accordi in buona fede, di dire la verità e di portare a termine ogni promessa fatta. Il trustee è concepito come una persona leale, che mantiene le promesse fatte e gli impegni presi.
- **Prevedibilità.** Questa credenza viene definita in termini di coerenza: le azioni di una persona sono coerenti a tal punto che è possibile prevedere quello che farà in una data situazione. Il trustee è concepito come una persona le cui azioni, coerenti e cognitivamente anticipabili, risultano in accordo con quanto previsto dal trustor.

La combinazione ottimale di queste quattro credenze porta a una solida intenzione a fidarsi e al successivo comportamento di fiducia. Nonostante questo, perché si provi fiducia non è necessario che tutte e quattro le credenze siano giudicate come possedute in misura elevata: in una specifica situazione, di solito, alcune credenze si rivelano più importanti di altre. Per esempio, in talune circostanze (come in un rapporto tra paziente e medico) il credere competente una persona è determinante ai fini della fiducia che vi si potrà riporre; diversamente, in altre circostanze (come in un rapporto tra manager e dipendente), altre credenze diventano preponderanti nel determinare il giudizio di fiducia, come ad esempio la benevolenza, che assicura che una certa informazione non finirà in mani disoneste.

Il presente studio si incentra sulle modalità attraverso cui si genera il giudizio di fiducia nelle persone. È verosimile ipotizzare che, a seconda delle situazioni, solo alcune delle competenze determineranno il giudizio di fiducia, e assumeranno importanza a scapito di altre. Appare quindi plausibile che le credenze potrebbero essere integrate nel concetto unitario di “affidabilità” attraverso un qualche processo di ponderazione. La presente ricerca è stata quindi realizzata con lo scopo di determinare le regole d'integrazione su cui si basa la valutazione multi-attributo dell'affidabilità.

5.2 Metodo

5.2.1 Partecipanti

Trentasei studenti dell'Università di Padova sono stati coinvolti nello studio (18 maschi e 18 femmine di età compresa fra 20 e 28 anni); essi sono stati assegnati in maniera casuale a una di tre distinte condizioni sperimentali. Quattro studenti hanno preso parte a una fase preparatoria pre-sperimentale.

5.2.2 Disegno sperimentale

Il disegno sperimentale era composto dai quattro attributi (benevolenza, competenza, onestà e prevedibilità¹), ognuno descritto attraverso uno di tre possibili livelli (basso, medio, alto). Al fine di analizzare il processo d'integrazione, gli attributi sono stati combinati a due a due, ottenendo sei disegni a due vie risultanti dalla combinazione a coppie dei sei attributi. Tale procedura è la stessa discussa in Anderson (1982, p. 92).

Il processo d'integrazione delle credenze è stato studiato in tre differenti contesti di fiducia: interpersonale, istituzionale e organizzativo. Dei 36 soggetti considerati, 12 sono stati assegnati al contesto interpersonale, 12 al contesto istituzionale e 12 al contesto organizzativo. Fondamentalmente, i contesti si distinguevano fra loro per la tipologia di situazioni che venivano presentate (si veda l'appendice C).

5.2.3 Apparato e stimoli

L'esperimento presentava 18 descrizioni di situazioni appartenenti ai 3 diversi contesti (6 interpersonali, 6 istituzionali e 6 organizzative: si veda l'appendice C). Per ogni situazione, all'intervistato era chiesto di quantificare il grado di fiducia che avrebbe riposto su una persona descritta da un particolare profilo di fiducia. Tali profili erano in totale 54, ed erano ottenuti

¹Nelle figure presentate più avanti, i quattro attributi sono sinteticamente indicati attraverso le lettere B per benevolenza, C per competenza, H per onestà e P per prevedibilità.

ITEM:	DEFINIZIONI	
<p>Stai partendo per le vacanze. Hai curato con tanta attenzione le piante nel tuo appartamento. Stai pensando che il tuo vicino di casa sarebbe disponibile ad innaffiarle. Devi però lasciargli le chiavi di casa. Quanta fiducia riponi nel tuo vicino sapendo che presenta le seguenti caratteristiche?</p>	<p>BENEVOLENZA L'interesse nei confronti del benessere di un'altra persona e la motivazione ad agire nel suo interesse.</p> <p>ONESTA' La capacità di prendere accordi in buona fede, di dire la verità e di portare a termine ogni promessa fatta.</p>	<p>COMPETENZA L'abilità di fare per un'altra persona ciò che egli ha bisogno che venga fatto.</p> <p>PREVEDIBILITA' Le azioni di una persona sono coerenti al punto che è possibile prevedere quello che farà in una data situazione.</p>
<p>Descrizione:</p> <p>1. bassa competenza e tanto benevolenza</p> <p>2. alta competenza e alta benevolenza</p> <p>3. media competenza e bassa benevolenza</p> <p>4. media competenza e media benevolenza</p> <p>5. bassa competenza e media benevolenza</p> <p>6. media competenza e alta benevolenza</p> <p>7. alta competenza e bassa benevolenza</p> <p>8. alta competenza e media benevolenza</p> <p>9. bassa competenza e bassa benevolenza</p>	<p>Totale sfiducia</p> <p>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</p> <p>Totale fiducia</p> <p>0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20</p>	<p>INDIETRO</p> <p>CONTINUA</p> <p>ESCI</p>

Figura 5.1: Interfaccia sperimentale. In alto a sinistra, l'item presenta la situazione, e di seguito si trovano i nove profili derivanti (in questo specifico caso) dalla combinazione degli attributi competenza e benevolenza. In alto a destra, sono riportate delle brevi descrizioni delle quattro credenze, per ricordare al rispondente la definizione di ognuna.

dai 6 disegni fattoriali (4 attributi presi a coppie) con $3 \times 3 = 9$ combinazioni di livelli (basso, medio, alto) ognuno. Quindi, i profili delineavano le caratteristiche della persona menzionata nella descrizione della situazione in termini di combinazione di attributi, dove ogni attributo possedeva uno di tre differenti livelli (ad es. un medico con alta onestà e bassa competenza). Accanto a ogni profilo era posta una scala di risposta a 21 punti: da 0 (totale sfiducia) a 20 (totale fiducia).

5.2.4 Procedura

I partecipanti lavoravano individualmente. A ognuno venivano presentate, a una a una, le sei situazioni inerenti il contesto assegnato. Il partecipante doveva leggere la descrizione della situazione e, per ogni singolo profilo, quantificare sulla scala di risposta il grado di fiducia, valutato secondo le proprie preferenze. I profili erano presentati in ordine casuale.

Ai partecipanti veniva proposta una preliminare fase di addestramento in modo che potessero familiarizzare col compito. Tale fase consisteva nella

lettura di una situazione (non inclusa nell’esperimento) con la richiesta di valutare i 54 profili ottenuti dalle combinazioni a coppie di tutti gli attributi.

Le analisi statistiche e grafiche sono state realizzate utilizzando il linguaggio R (R Development Core Team, 2011).

5.3 Risultati

In primo luogo sono state analizzate le risposte dei quattro soggetti che hanno partecipato alla fase preliminare della ricerca, al fine di valutare la plausibilità dei modelli d’integrazione. Dalle peculiarità delle configurazioni delle risposte è stata riscontrata un’evidenza a favore del modello averaging.

Per i 36 soggetti che hanno preso parte alla fase sperimentale, sono state condotte in prima battuta analisi separate sui dati di ognuno. Per ogni partecipante sono stati stimati i parametri del modello averaging attraverso il metodo R-Average, utilizzando l’implementazione *rAverage*, versione 0.4-0. Lo stato iniziale è stato eliminato dal modello per ottenere massima unicità dei parametri relativi agli stimoli manipolati (Anderson, 1982; Noventa et al. 2010).

Un partecipante appartenente al gruppo che valutava le situazioni del contesto organizzativo è stato escluso dalle analisi a causa delle anomalie nelle risposte; dunque, per questo gruppo sono stati considerati solo 11 soggetti. I restanti 35 partecipanti mostravano delle risposte coerenti senza evidenziare particolari anomalie.

La tabella 5.1 riporta gli indici R^2 , AIC e BIC per l’analisi sui 35 partecipanti, e la figura 5.2 compara l’indice R^2 di ogni singolo partecipante per il miglior modello averaging selezionato contro quello stimato per il caso a pesi uguali. La maggior parte dei soggetti mostra degli indici che indicano in generale un buon adattamento ai dati, con valori di scala ben differenziati e pesi parzialmente uguali entro e tra fattori. Solo per il partecipante 7 nel gruppo “contesto interpersonale”, per i partecipanti 4 e 11 nel gruppo “contesto istituzionale”, e per i partecipanti 5 e 11 nel gruppo “contesto organizzativo”, è stato identificato un modello averaging a pesi completamente diversi. Al contrario, solo per il soggetto 2 nel contesto organizzativo è stato

(a) Contesto interpersonale

ID	1	2	3	4	5	6	7*	8	9	10	11	12
R^2	0.76	0.85	0.86	0.85	0.81	0.82	0.80	0.72	0.73	0.88	0.85	0.79
AIC	655	499	386	501	459	553	649	506	579	463	542	545
BIC	723	559	442	573	526	613	721	561	638	522	605	609

(b) Contesto istituzionale

ID	1	2	3	4*	5	6	7	8	9	10	11*	12
R^2	0.82	0.82	0.85	0.86	0.85	0.83	0.91	0.92	0.77	0.60	0.52	0.83
AIC	712	644	193	506	360	674	435	373	651	895	428	613
BIC	780	708	261	578	437	742	503	446	719	950	504	672

(c) Contesto organizzativo

ID	1	2**	3	4	5*	6	7	8	9	10	11*
R^2	0.86	0.75	0.73	0.91	0.92	0.91	0.89	0.91	0.88	0.55	0.92
AIC	534	548	689	447	334	357	367	415	347	863	288
BIC	602	595	757	520	406	434	439	478	406	923	347

Tabella 5.1: Indici di adattamento per i singoli partecipanti (identificati da un codice numerico indicato alla voce ID) per i tre contesti. (*) Modello stimato a pesi completamente diversi; (**) modello stimato a pesi completamente uguali.

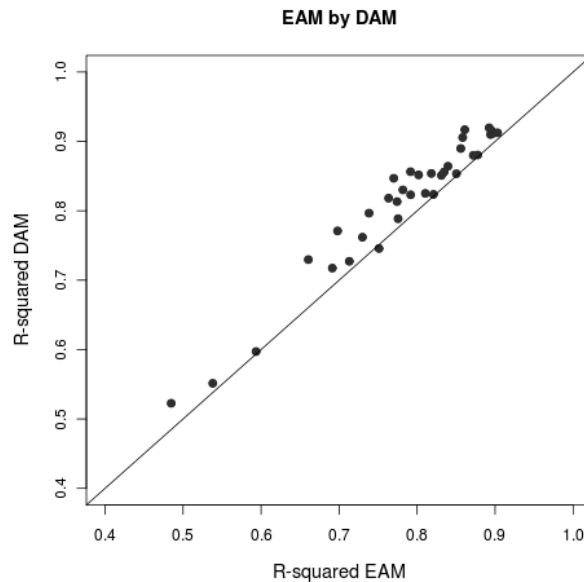


Figura 5.2: Comparazione dell'indice R^2 tra i modelli averaging a pesi uguali e diversi.

selezionato un modello a pesi completamente uguali. Sono stati osservati anche alcuni casi di modello semilineare (Anderson, 1982): i partecipanti 4, 6, e da 9 a 12 nel contesto interpersonale, i partecipanti 4, 7, 8, e 10 nel contesto istituzionale, e i partecipanti 4, 8, e 10 nel contesto organizzativo.

L'analisi di cluster è stata applicata per analizzare le prossimità fra i soggetti (Struyf et al., 1997). Dato che i risultati mostravano delle similarità delle valutazioni per la maggior parte dei partecipanti, questi sono stati accorpati e sono state eseguite due separate analisi della varianza utilizzando come variabile dipendente uno dei due parametri stimati attraverso R-Average (i parametri s oppure i parametri t) e come variabili indipendenti il contesto (interpersonale, istituzionale, organizzativo), l'attributo (benevolenza, competenza, onestà, prevedibilità) e il livello (basso, medio, alto).

Per quanto riguarda i valori di scala s , come atteso, è stata rilevata una differenza statisticamente significativa negli effetti principali per i fattori Livello ($F_{(2,64)} = 677.22, p < 0.001, \eta^2 = 0.95$) e Attributo ($F_{(3,96)} = 11.04, p < 0.001, \eta^2 = 0.26$). Inoltre, sono state rilevate due interazioni significative: Livello \times Attributo ($F_{(6,192)} = 19.33; p < 0.001, \eta^2 = 0.38$) e Livello \times Contesto ($F_{(12,192)} = 1.94, p < 0.05, \eta^2 = 0.11$). Le stesse significatività statistiche sono state trovate per i pesi t stimati, negli effetti principali sia per il fattore Livello ($F_{(2,64)} = 39.40, p < 0.001, \eta^2 = 0.55$) che per il fattore Attributo ($F_{(3,96)} = 28.24, p < 0.001, \eta^2 = 0.49$) e nell'interazione Livello \times Attributo ($F_{(6,192)} = 2.52, p < 0.05, \eta^2 = 0.07$). La figura 5.3 mostra i grafici per le interazioni Livello \times Attributo per i valori di scala (grafico di sinistra) e per i pesi (grafico di destra).

5.4 Discussione

I risultati presentati conducono a importanti conclusioni. Secondo McKnight et al. (1998), le quattro principali credenze giocano un ruolo fondamentale nel giudizio di fiducia. Inoltre, in accordo con la teoria dell'integrazione delle informazioni e la misurazione funzionale, un modello averaging sembra essere in grado di spiegare le risposte individuali. Mentre la maggior parte dei partecipanti (34) può essere ricondotta a un caso a pesi diversi, solo un

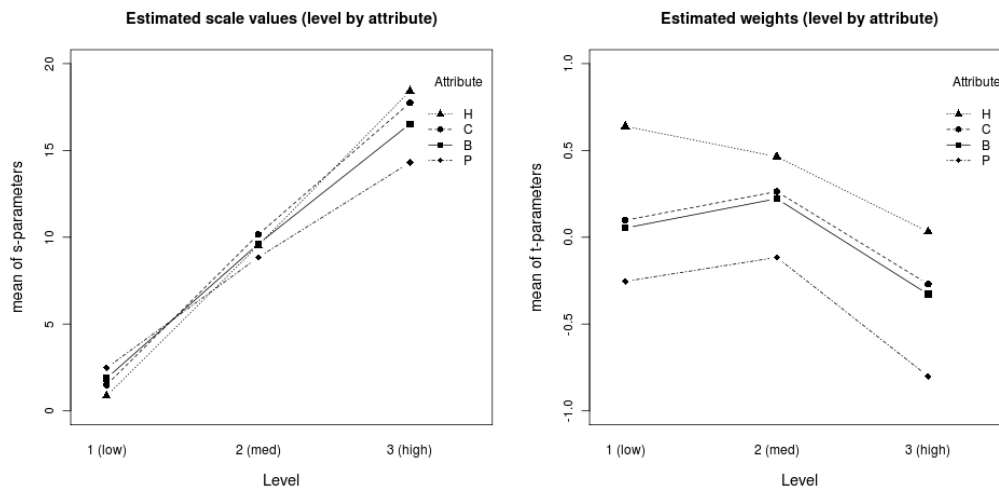


Figura 5.3: Medie dei parametri stimati in funzione dei tre livelli (basso, medio e alto) e dei quattro attributi (onestà, competenza, benevolenza e prevedibilità). Il grafico di sinistra riporta le medie dei valori di scala s e quello di destra le medie dei pesi t ; per semplicità di lettura, le barre d'errore non sono state riportate.

partecipante mostra un chiaro pattern che si conforma a un modello a pesi uguali.

I valori di scala mostrano un netto andamento lineare con inclinazione maggiore per onestà e competenza (figura 5.3, grafico di sinistra), e anche i pesi mostrano valori medi più elevati per onestà e competenza (figura 5.3, grafico di destra). Questi risultati sono coerenti con l'idea che attributi diversi possano influenzare diversamente il giudizio finale: onestà e competenza sembrano preponderanti, mentre la prevedibilità appare meno rilevante.

Un'altra interessante conclusione riguarda l'onestà, il cui basso livello presenta un peso superiore rispetto agli altri livelli, per cui la bassa onestà sembra giocare un importante ruolo nel determinare un giudizio finale di polarità negativa. Inoltre, la differente inclinazione dell'andamento dei livelli degli attributi nei tre contesti (figura 5.4) suggerisce un ruolo preminente dell'onestà nel contesto interpersonale e della competenza nel contesto istituzionale.

In conclusione, la teoria dell'integrazione delle informazioni e la misurazione funzionale sembrano rappresentare un interessante approccio per comprendere le credenze di fiducia e il loro ruolo all'interno della generale teoria

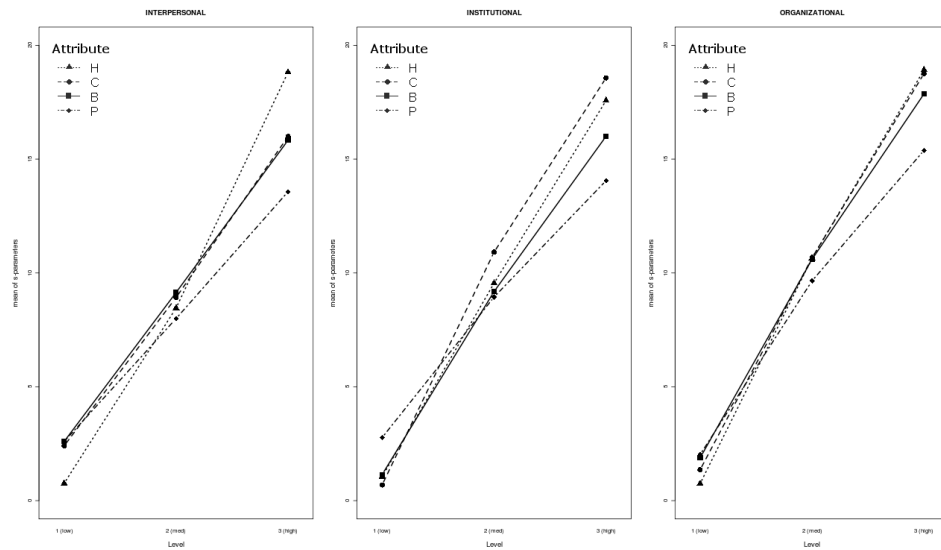


Figura 5.4: Medie dei valori di scala stimati in funzione dei tre livelli (basso, medio e alto), dei quattro attributi (onestà, competenza, benevolenza e prevedibilità) nei tre diversi contesti situazionali (interpersonale, istituzionale, organizzativo). Per semplicità di lettura, le barre d'errore non sono state riportate.

della fiducia. Il modello averaging a pesi diversi può quindi diventare la chiave per la comprensione di tali fenomeni. Nonostante questa famiglia di modelli sia apparsa in letteratura (Anderson, 1991) meno rispetto ad altre, nasconde un grande potenziale che può essere sfruttato negli studi futuri per analizzare il ruolo di variabili critiche che altri modelli più semplici non sono in grado di considerare.

Lo studio condotto presenta diversi limiti, e la generalizzabilità dei risultati dev'essere considerata con cautela. I dati si riferiscono a un gruppo di studenti universitari italiani; sarebbe importante verificare se i processi d'integrazione individuati e l'importanza degli attributi stimata corrisponda a un generico modello con validità cross-culturale. Inoltre, lo studio soffre di una limitata validità ecologica: nella vita reale, gli stessi soggetti potrebbero rispondere diversamente. Sono quindi necessari nuovi studi per confermare i dati qui discussi.

Conclusione

Il presente lavoro di ricerca si è proposto di sviluppare e applicare un nuovo metodo per la stima dei parametri del modello averaging della teoria dell'integrazione delle informazioni. Tale lavoro risponde a una sentita necessità della comunità scientifica che opera nell'ambito della misurazione funzionale. Il modello averaging, infatti, presenta un grosso potenziale che fino a oggi non è stato sfruttato appieno, proprio a causa della carenza di un solido impianto statistico che ne permetta la completa applicazione.

Secondo Box e Draper (1987) sostanzialmente tutti i modelli sono sbagliati, ma alcuni si possono rivelare utili. In quest'ottica, quindi, il problema non è capire se un modello sia giusto o sbagliato, ma è capire quanto sbagliato sia. Pertanto, sulla base dell'errore che il modello contiene, si dovrebbe stabilire se lo si possa o meno considerare realmente utile per sintetizzare le informazioni contenute nei dati e trarne delle inferenze. Seguendo questa logica, il modello averaging non dovrebbe essere considerato come una legge universale che regola il giudizio umano, ma dovrebbe essere concepito e utilizzato come una delle possibili modalità attraverso le quali la percezione e il giudizio umano possono essere descritti e sintetizzati. Quella averaging sarebbe quindi una delle possibili regole attraverso cui il prodotto dell'integrazione può essere scisso. Rispetto ad altre regole essa comporta diversi vantaggi, in quanto permette di descrivere in maniera semplice ma efficace complesse interazioni fra attributi; tuttavia, rimane un'approssimazione, e come tale dovrebbe essere trattata.

Il metodo R-Average (cap. 2) si è mostrato un efficace strumento per la stima dei parametri e la selezione del modello più adeguato a descrivere i dati. Questo strumento mette in grado lo sperimentatore di selezionare il modello

valutando la plausibilità di una serie di possibili alternative, stabilendo quale fra queste apporta il maggior grado di spiegazione del fenomeno sotto analisi mantenendo la minima complessità strutturale possibile.

Tale metodologia permette alla misurazione funzionale di compiere un passo avanti e apre nuovi scenari nelle possibilità applicative della teoria dell'integrazione delle informazioni. Il metodo appare ormai maturo, ma la sua applicazione richiede comunque qualche cautela: se sotto le più comuni condizioni di stima le procedure convergono verso ottimi risultati, in alcuni casi si evidenziano delle difficoltà che nel prossimo futuro dovranno essere affrontate e necessariamente superate. Tuttavia, anche al momento attuale, la soluzione è già in mano agli sperimentatori. Ottenere più misurazioni su uno stesso soggetto in una medesima condizione sperimentale si può rivelare di vitale importanza per oltrepassare tali problematiche, e già questo può risultare sufficiente per eludere le criticità insite nella natura del modello. Inoltre, lo studio preliminare del comportamento dei singoli soggetti è un punto di partenza necessario per passare solo in un secondo momento ad analisi sovraordinate, nelle quali si potranno creare gruppi di soggetti per ottenere risultati complessivi.

Il problema del raggruppamento dei soggetti sulla base della somiglianza delle loro risposte sarà uno dei prossimi capitoli che la ricerca in misurazione funzionale dovrà affrontare. Difatti, se da un lato le analisi sul singolo soggetto da sole risultano troppo dispersive in termini informativi, le analisi di gruppo potrebbero condurre alla perdita di importanti informazioni, mascherando dei fenomeni che invece si evidenzerebbero se il raggruppamento fosse realizzato secondo una logica e una metodologia studiate *ad hoc*.

Determinati fenomeni, come quelli analizzati nel presente lavoro (capp. 3-5), ora possono essere compresi con maggior dettaglio. Pertanto, si aprono nuovi scenari nell'ambito della teoria dell'integrazione delle informazioni, e se lo sviluppo delle procedure di stima che compongono R-Average terrà il passo con le necessità applicative, il modello averaging e la misurazione funzionale potranno ambire a nuove prospettive.

Appendice A

Risultati degli studi Monte Carlo

Trials	σ_ϵ	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	w_{A_1}	w_{A_2}	w_{A_3}	w_{B_1}	w_{B_2}	w_{B_3}	R^2	
		5	10	15	5	10	15	4	1	4	1	3	3		
5	0.5	Mdn	4.99	10.17	15.11	4.96	10.03	15.18	3.88	1.01	3.88	1.00	2.85	2.79	
		Iqr	0.28	0.48	0.33	0.45	0.26	0.46	1.44	0.48	1.64	0.50	1.24	1.09	0.98
	1.0	Mdn	4.99	10.49	15.24	4.98	10.12	15.59	3.54	1.00	3.54	1.00	2.50	2.33	0.94
		Iqr	0.53	1.08	0.71	1.18	0.57	1.21	2.21	0.92	3.98	1.06	1.52	1.36	
	1.5	Mdn	4.97	10.69	15.47	5.12	10.20	16.00	3.09	1.00	3.09	1.39	2.09	1.98	0.87
		Iqr	0.98	1.44	1.19	1.75	0.99	2.32	2.72	2.21	3.70	1.81	1.74	1.74	
10	0.5	Mdn	5.00	10.14	15.07	5.03	10.03	15.13	3.83	1.00	3.83	1.00	2.83	2.81	0.98
		Iqr	0.19	0.34	0.25	0.31	0.18	0.32	2.19	0.63	2.64	0.61	1.82	1.72	
	1.0	Mdn	5.00	10.29	15.11	4.99	10.06	15.34	3.88	1.03	3.88	1.00	2.73	2.59	0.93
		Iqr	0.39	0.78	0.46	0.70	0.38	0.76	2.06	0.64	2.89	0.63	1.67	1.39	
	1.5	Mdn	4.98	10.54	15.23	4.92	10.10	15.62	3.60	1.02	3.60	1.00	2.55	2.38	0.86
		Iqr	0.69	1.23	0.79	1.44	0.61	1.40	2.39	0.84	3.72	0.94	1.91	1.54	
15	0.5	Mdn	5.00	10.10	15.05	5.03	10.02	15.08	3.88	1.01	3.88	1.00	2.88	2.87	0.98
		Iqr	0.15	0.26	0.19	0.24	0.14	0.22	2.29	0.65	2.71	0.60	1.96	1.82	
	1.0	Mdn	5.00	10.22	15.10	5.02	10.04	15.24	3.76	1.02	3.76	1.00	2.73	2.65	0.93
		Iqr	0.36	0.60	0.37	0.58	0.29	0.55	1.68	0.58	2.54	0.56	1.69	1.37	
	1.5	Mdn	4.96	10.42	15.17	4.95	10.07	15.46	3.71	1.02	3.71	1.00	2.68	2.48	0.86
		Iqr	0.52	1.00	0.61	1.04	0.46	1.08	1.68	0.64	3.34	0.70	1.65	1.32	

Tabella A.1: Risultati delle simulazioni Monte Carlo per il primo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	s_0	w_0	s_{A_1}			s_{A_2}			s_{A_3}			s_{B_1}			s_{B_2}			s_{B_3}			w_{A_1}	w_{A_2}	w_{A_3}	w_{B_1}	w_{B_2}	w_{B_3}	R^2				
				5	10	15	5	10	15	5	10	15	4	1	4	1	4	1	4	1	4								1	4	1	4
5	0.5	Mdn	1.91	1.04	5.05	9.91	14.92	5.05	10.15	15.32	4.16	1.08	4.16	1.00	2.82	2.81	0.98	Iqr	2.14	0.51	0.31	1.26	0.96	0.99	0.65	1.55	0.90	0.45	1.08	0.51	0.46	0.47
		Mdn	1.05	1.02	5.06	9.99	14.83	4.98	10.19	15.56	3.91	1.00	3.91	1.07	2.69	2.64	0.92	Iqr	3.05	0.77	0.65	2.85	1.76	1.57	1.44	2.63	1.34	0.88	1.99	1.79	0.88	0.80
	Mdn	0.31	1.01	5.05	9.33	15.27	5.09	10.45	16.09	3.43	1.00	3.43	1.67	2.37	2.42	0.84	Iqr	3.24	1.03	0.93	3.99	2.56	1.89	2.24	4.20	1.63	3.37	1.84	2.19	1.19	1.21	
	Mdn	1.93	1.00	5.03	9.89	14.96	5.00	10.09	15.23	4.06	1.04	4.06	1.00	2.88	2.86	0.98	Iqr	1.52	0.38	0.26	0.83	0.68	0.75	0.44	0.94	0.82	0.33	0.82	0.37	0.54	0.45	
	Mdn	1.63	1.02	5.07	9.96	14.84	5.02	10.17	15.50	4.00	1.05	4.00	1.00	2.71	2.66	0.92	Iqr	2.96	0.61	0.45	1.84	1.15	1.32	0.91	2.05	1.10	0.61	1.17	0.76	0.58	0.57	
	Mdn	0.74	1.00	5.08	9.84	14.81	5.11	10.28	15.78	3.81	1.00	3.81	1.11	2.54	2.48	0.83	Iqr	3.01	0.75	0.66	2.76	1.79	1.59	1.43	3.18	1.33	1.12	1.90	2.04	0.89	0.84	
10	0.5	Mdn	2.01	1.01	5.02	9.93	14.95	5.01	10.08	15.18	4.01	1.03	4.01	1.00	2.91	2.92	0.98	Iqr	1.23	0.33	0.20	0.67	0.49	0.64	0.37	0.78	0.83	0.29	0.75	0.30	0.56	0.56
		Mdn	1.95	1.01	5.01	9.88	14.91	4.94	10.13	15.39	4.07	1.05	4.07	1.00	2.81	2.78	0.91	Iqr	2.82	0.83	0.38	1.44	1.01	1.21	0.82	1.77	1.37	0.68	1.34	0.77	0.71	0.69
	Mdn	1.40	0.98	5.04	9.83	14.87	5.05	10.26	15.58	3.87	1.03	3.87	1.00	2.56	2.56	0.83	Iqr	3.25	1.04	0.56	1.97	1.46	1.38	1.24	2.58	1.55	1.00	2.31	1.24	0.92	0.97	
	Mdn	1.63	1.02	5.07	9.96	14.84	5.02	10.17	15.50	4.00	1.05	4.00	1.00	2.71	2.66	0.92	Iqr	2.96	0.61	0.45	1.84	1.15	1.32	0.91	2.05	1.10	0.61	1.17	0.76	0.58	0.57	
	Mdn	0.74	1.00	5.08	9.84	14.81	5.11	10.28	15.78	3.81	1.00	3.81	1.11	2.54	2.48	0.83	Iqr	3.01	0.75	0.66	2.76	1.79	1.59	1.43	3.18	1.33	1.12	1.90	2.04	0.89	0.84	
	Mdn	2.01	1.01	5.02	9.93	14.95	5.01	10.08	15.18	4.01	1.03	4.01	1.00	2.91	2.92	0.98	Iqr	1.23	0.33	0.20	0.67	0.49	0.64	0.37	0.78	0.83	0.29	0.75	0.30	0.56	0.56	

Tabella A.2: Risultati delle simulazioni Monte Carlo per il secondo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	w_{A_1}	w_{A_2}	w_{A_3}	w_{B_1}	w_{B_2}	w_{B_3}	R^2	
		5	10	15	5	10	15	4	1	4	2	2	2		
5	0.5	Mdn	4.93	10.07	15.12	5.00	10.04	15.10	3.99	1.00	3.99	2.07	2.08	2.04	0.98
		Iqr	0.32	0.37	0.33	0.35	0.27	0.47	1.57	0.49	1.68	0.71	0.70	0.76	
	1.0	Mdn	4.80	10.19	15.35	4.97	10.06	15.30	3.63	1.00	3.63	2.09	2.11	2.05	0.94
		Iqr	0.62	0.79	0.64	0.77	0.55	1.11	1.53	0.83	1.35	1.04	1.08	1.18	
	1.5	Mdn	4.69	10.30	15.50	4.81	10.17	15.60	3.55	1.00	3.55	2.10	2.12	2.06	0.87
		Iqr	1.03	1.15	1.15	1.27	0.93	1.80	1.97	2.44	2.06	1.43	1.43	1.45	
10	0.5	Mdn	4.96	10.05	15.06	4.98	10.02	15.07	4.16	1.00	4.16	2.09	2.09	2.08	0.98
		Iqr	0.20	0.27	0.22	0.27	0.20	0.35	1.52	0.39	1.51	0.68	0.66	0.70	
	1.0	Mdn	4.87	10.15	15.22	5.00	10.06	15.13	3.82	1.00	3.82	2.12	2.11	2.05	0.93
		Iqr	0.44	0.58	0.43	0.50	0.37	0.70	1.52	0.70	1.50	1.12	1.12	1.19	
	1.5	Mdn	4.77	10.21	15.41	4.99	10.14	15.27	3.57	1.00	3.57	2.11	2.12	2.04	0.86
		Iqr	0.66	0.87	0.73	0.77	0.59	1.17	2.38	0.95	2.43	1.55	1.55	1.64	
15	0.5	Mdn	4.98	10.05	15.05	4.99	10.03	15.05	4.15	1.00	4.15	2.07	2.08	2.06	0.98
		Iqr	0.16	0.23	0.16	0.19	0.16	0.25	1.49	0.39	1.41	0.70	0.68	0.72	
	1.0	Mdn	4.92	10.11	15.17	5.00	10.04	15.11	3.84	1.00	3.84	2.06	2.07	2.02	0.93
		Iqr	0.38	0.43	0.36	0.37	0.31	0.53	1.92	0.67	2.03	1.29	1.30	1.39	
	1.5	Mdn	4.81	10.17	15.29	5.03	10.08	15.18	3.59	1.00	3.59	2.06	2.06	2.01	0.86
		Iqr	0.51	0.64	0.56	0.64	0.51	0.89	2.53	0.99	2.45	1.63	1.71	1.75	

Tabella A.3: Risultati delle simulazioni Monte Carlo per il terzo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	s_0	w_0	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	w_{A_1}	w_{A_2}	w_{A_3}	w_{B_1}	w_{B_2}	w_{B_3}	R^2	
																	1
5	0.5	Mdn	2.01	1.06	4.96	10.22	15.22	4.99	9.92	14.79	4.09	1.00	4.09	2.20	2.22	2.23	0.98
		Iqr	2.59	0.99	0.36	1.56	15.22	0.57	0.73	1.29	2.27	0.68	2.65	0.60	0.55	0.57	
	1.0	Mdn	2.02	1.21	4.86	9.83	15.98	5.01	9.97	14.85	3.88	1.00	3.88	2.42	2.45	2.46	0.91
		Iqr	3.66	1.84	0.76	4.00	2.34	1.07	1.43	2.85	1.32	1.94	1.23	1.08	1.10	1.10	
	1.5	Mdn	1.96	0.58	4.79	8.75	16.28	4.98	10.12	15.38	1.56	1.56	1.56	1.00	1.02	1.03	0.82
		Iqr	4.03	2.20	1.11	2.88	3.20	1.59	2.24	4.20	1.55	4.98	1.56	1.40	1.39	1.41	
10	0.5	Mdn	2.09	1.09	4.99	10.25	15.02	4.97	9.90	14.87	4.35	1.00	4.35	2.22	2.24	2.22	0.97
		Iqr	1.70	0.81	0.26	1.09	0.80	0.43	0.58	1.17	1.23	0.57	1.32	0.36	0.35	0.36	
	1.0	Mdn	2.10	1.10	4.89	10.13	15.70	4.99	9.88	14.70	3.89	1.00	3.89	2.38	2.40	2.40	0.91
		Iqr	3.42	1.27	0.52	2.12	1.65	0.75	0.99	1.83	1.17	0.94	0.97	0.78	0.78	0.79	
	1.5	Mdn	2.19	1.30	4.87	9.75	16.00	4.98	10.01	14.85	4.04	1.00	4.04	2.57	2.57	2.61	0.81
		Iqr	3.89	1.79	0.82	4.02	2.35	1.06	1.50	2.84	1.26	3.56	1.18	1.07	1.04	1.06	
15	0.5	Mdn	2.16	1.11	4.97	10.20	15.05	4.97	9.92	14.94	4.33	1.00	4.33	2.21	2.21	2.21	0.97
		Iqr	1.39	0.69	0.21	0.87	0.65	0.36	0.48	0.99	1.00	0.47	1.03	0.30	0.31	0.31	
	1.0	Mdn	2.02	1.08	4.93	10.24	15.48	5.02	9.85	14.60	3.96	1.00	3.96	2.37	2.42	2.41	0.90
		Iqr	2.84	1.04	0.43	1.98	1.38	0.57	0.82	1.38	1.76	0.80	1.77	0.67	0.64	0.64	
	1.5	Mdn	1.92	1.16	4.87	10.13	15.78	5.03	9.84	14.68	3.96	1.00	3.96	2.56	2.58	2.58	0.81
		Iqr	3.61	1.52	0.66	3.27	1.96	0.86	1.21	2.20	0.95	1.20	0.90	0.91	0.88	0.88	

Tabella A.4: Risultati delle simulazioni Monte Carlo per il quarto gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	$w_0 < 0.1$	
		Group 1	Group 3
5	0.5	933	969
	1.0	749	821
	1.5	634	691
10	0.5	976	991
	1.0	852	913
	1.5	697	818
15	0.5	999	999
	1.0	901	951
	1.5	634	691

Tabella A.5: Numero di modelli in cui il parametro w_0 è stimato inferiore a 0.1, per i due gruppi di parametri in cui lo stato iniziale è nullo.

Trials	σ_ϵ	Group 1			Group 2			Group 3			Group 4		
		EAM	DAM	IC	EAM	DAM	IC	EAM	DAM	IC	EAM	DAM	IC
5	0.5	0	124	876	0	75	925	0	33	967	0	40	960
	1.0	8	42	950	19	43	938	24	10	966	195	17	788
	1.5	169	25	806	252	15	733	267	17	716	521	14	465
10	0.5	0	119	881	0	76	924	0	55	945	0	54	946
	1.0	0	58	942	0	38	962	0	18	982	16	32	952
	1.5	10	29	961	33	33	934	37	11	952	226	10	764
15	0.5	0	156	844	0	93	907	0	80	920	0	57	943
	1.0	0	69	931	0	100	900	0	31	969	0	24	976
	1.5	0	40	960	3	65	932	3	19	978	95	15	890

Tabella A.6: Numero di modelli, stimati da ogni procedura, selezionati come migliore dalla comparazione finale, per ogni simulazione Monte Carlo.

Trials	σ_ϵ	s_0	w_0	s_{A_i}			s_{B_i}			w_{A_i}			w_{B_i}			R^2
				s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	w_{A_1}	w_{A_2}	w_{A_3}	w_{B_1}	w_{B_2}	w_{B_3}	
5	0.5	Mdn	2.01	1.07	5.05	9.99	14.88	5.03	10.16	15.35	4.28	1.06	4.28	1.00	2.84	0.98
		Iqr	2.13	0.53	0.31	1.34	0.90	0.99	0.64	1.50	0.88	0.44	0.78	0.50	0.42	
	Mdn	1.08	1.06	5.05	10.09	14.82	4.99	10.20	15.56	4.09	1.00	4.09	1.08	2.79	2.75	0.92
	Iqr	3.09	0.78	0.66	3.01	1.72	1.58	1.44	2.62	1.25	0.90	1.86	1.81	0.86	0.80	
	Mdn	0.47	1.06	5.04	9.39	15.29	5.09	10.42	16.19	3.54	1.00	3.54	1.68	2.41	2.45	0.84
	Iqr	3.28	1.02	0.92	4.27	2.59	1.87	2.22	4.24	1.72	3.40	1.85	2.16	1.21	1.24	
10	0.5	Mdn	1.98	1.01	5.03	9.97	14.91	4.98	10.09	15.28	4.09	1.02	4.09	1.00	2.84	0.98
		Iqr	1.48	0.41	0.25	0.91	0.69	0.77	0.43	1.01	0.85	0.32	0.56	0.38	0.32	
	Mdn	1.77	1.05	5.06	10.00	14.84	4.99	10.17	15.53	4.11	1.03	4.11	1.00	2.72	2.69	0.92
	Iqr	3.03	0.60	0.46	1.90	1.12	1.33	0.92	2.00	0.98	0.60	1.05	0.77	0.57	0.56	
	Mdn	0.73	1.02	5.09	9.92	14.82	5.14	10.29	15.79	3.89	1.00	3.89	1.13	2.56	2.53	0.83
	Iqr	3.13	0.72	0.66	2.80	1.78	1.57	1.42	3.05	1.31	1.04	1.65	2.02	0.91	0.86	
15	0.5	Mdn	2.03	1.02	5.03	10.01	14.89	4.98	10.08	15.24	4.04	1.00	4.04	1.00	2.91	0.98
		Iqr	1.17	0.32	0.19	0.69	0.55	0.61	0.38	0.80	0.89	0.25	0.60	0.29	0.32	
	Mdn	1.95	1.05	5.03	9.94	14.84	4.93	10.14	15.43	4.24	1.05	4.24	1.00	2.83	2.79	0.91
	Iqr	2.65	0.58	0.38	1.43	0.99	1.22	0.78	1.77	0.83	0.50	0.72	0.62	0.50	0.48	
	Mdn	1.30	0.97	5.08	9.93	14.67	5.05	10.30	15.70	3.85	1.00	3.85	1.02	2.49	2.47	0.83
	Iqr	3.12	0.65	0.57	2.13	1.43	1.36	1.22	2.61	0.96	0.80	1.29	1.38	0.71	0.71	

Tabella A.7: Risultati delle simulazioni Monte Carlo per il secondo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. A differenza della tabella A.2, in questo caso la stima è realizzata utilizzando la versione ridotta della procedura Information Criteria. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	s_0	w_0	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	w_{A_1}	w_{A_2}	w_{A_3}	w_{B_1}	w_{B_2}	w_{B_3}	R^2	
																	1
5	0.5	Mdn	2.01	1.05	4.96	10.13	15.22	4.99	9.95	14.76	4.03	1.00	4.03	2.11	2.20	2.21	0.98
		Iqr	2.63	0.40	0.36	1.32	1.38	0.58	0.74	1.36	1.69	0.60	1.93	0.93	0.88	0.88	
5	1.0	Mdn	1.98	1.23	4.88	9.88	15.92	4.99	9.98	14.94	3.98	1.00	3.98	2.50	2.52	2.53	0.91
		Iqr	3.67	0.66	0.72	3.98	2.34	1.07	1.46	2.71	1.66	1.25	1.72	1.30	1.35	1.31	
5	1.5	Mdn	1.85	0.56	4.80	8.78	16.26	4.99	10.13	15.38	1.56	1.56	1.56	1.00	1.01	1.02	0.82
		Iqr	3.91	1.05	1.06	3.64	3.20	1.60	2.29	4.24	1.98	3.44	2.01	1.52	1.50	1.52	
10	0.5	Mdn	1.99	1.02	5.00	10.05	15.04	5.00	9.96	14.91	4.08	1.00	4.08	2.03	2.07	2.07	0.97
		Iqr	1.78	0.31	0.27	0.90	0.79	0.44	0.56	1.11	1.34	0.45	1.63	0.72	0.63	0.67	
10	1.0	Mdn	2.03	1.08	4.89	10.14	15.69	4.99	9.84	14.62	3.84	1.00	3.84	2.42	2.43	2.41	0.91
		Iqr	3.43	0.47	0.54	2.15	1.70	0.73	0.95	1.84	1.42	0.78	1.50	1.06	1.11	1.14	
10	1.5	Mdn	2.05	1.27	4.88	9.83	15.92	4.98	9.97	14.79	4.14	1.00	4.14	2.64	2.63	2.66	0.81
		Iqr	3.86	0.78	0.80	4.04	2.38	1.06	1.56	2.81	1.69	1.57	1.81	1.30	1.33	1.31	
15	0.5	Mdn	2.05	1.02	4.99	10.01	15.07	4.99	9.99	15.01	4.03	1.00	4.03	1.99	2.01	2.01	0.97
		Iqr	1.38	0.27	0.21	0.72	0.70	0.36	0.45	0.92	1.06	0.37	1.26	0.59	0.54	0.56	
15	1.0	Mdn	1.94	1.06	4.95	10.19	15.47	5.02	9.86	14.62	3.90	1.00	3.90	2.36	2.38	2.36	0.90
		Iqr	2.79	0.40	0.44	1.82	1.36	0.55	0.81	1.42	1.19	0.67	1.36	0.99	0.95	0.98	
15	1.5	Mdn	1.93	1.15	4.88	10.21	15.77	5.04	9.84	14.68	3.98	1.00	3.98	2.53	2.54	2.54	0.81
		Iqr	3.61	0.54	0.64	3.15	2.02	0.86	1.21	2.12	1.55	0.96	1.61	1.21	1.22	1.24	

Tabella A.8: Risultati delle simulazioni Monte Carlo per il quarto gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. A differenza della tabella A.4, in questo caso la stima è realizzata utilizzando la versione ridotta della procedura Information Criteria. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	t_{A_1}	t_{A_2}	t_{A_3}	t_{B_1}	t_{B_2}	t_{B_3}	R^2	
		Mdn	Iqr	Mdn	Iqr	Mdn	Iqr	Mdn	Iqr	Mdn	Iqr	Mdn	Iqr		
5	0.5	5	10	15	5	10	15	0.56	-0.83	0.56	-0.83	0.27	0.27	0.98	
		4.92	10.16	15.19	4.54	10.05	15.26	0.57	-0.80	0.57	-0.87	0.30	0.26		
	0.28	0.52	0.34	0.79	0.27	0.44	0.21	0.22	0.24	0.25	0.18	0.24			
	Mdn	4.80	10.69	15.54	3.94	10.16	15.82	0.57	-0.78	0.59	-0.89	0.34	0.22		0.94
		0.67	1.53	1.10	4.87	0.62	1.99	0.48	0.50	0.55	0.63	0.39	0.69		
	Mdn	4.60	10.82	15.82	3.34	10.26	17.85	0.28	-0.34	0.43	-0.47	0.00	-0.14		0.86
2.23		1.48	1.91	4.17	1.14	4.29	0.71	0.97	0.99	0.86	0.84	0.69			
10	0.5	5	10	15	5	10	15	0.57	-0.80	0.56	-0.87	0.29	0.26	0.98	
		4.93	10.12	15.15	4.62	10.04	15.23	0.57	-0.80	0.56	-0.87	0.29	0.26		
	0.19	0.32	0.24	0.50	0.18	0.28	0.14	0.14	0.16	0.16	0.13	0.15			
	Mdn	4.88	10.29	15.26	4.43	10.08	15.41	0.59	-0.79	0.57	-0.91	0.32	0.26		0.93
		0.40	0.86	0.54	1.62	0.39	0.77	0.31	0.32	0.33	0.37	0.24	0.35		
	Mdn	4.78	10.75	15.60	3.16	10.15	15.9	0.59	-0.78	0.58	-0.95	0.36	0.25		0.86
0.82		1.64	1.17	4.81	0.69	2.3	0.57	0.60	0.59	0.72	0.47	0.75			
15	0.5	5	10	15	5	10	15	0.57	-0.80	0.56	-0.87	0.28	0.27	0.98	
		4.93	10.09	15.13	4.61	10.03	15.19	0.57	-0.80	0.56	-0.87	0.28	0.27		
	0.15	0.25	0.19	0.38	0.15	0.22	0.11	0.12	0.12	0.13	0.09	0.12			
	Mdn	4.91	10.19	15.22	4.52	10.06	15.31	0.57	-0.79	0.56	-0.87	0.30	0.25		0.93
		0.34	0.57	0.40	0.98	0.29	0.52	0.25	0.26	0.27	0.31	0.20	0.29		
	Mdn	4.81	10.53	15.43	3.96	10.11	15.68	0.58	-0.79	0.61	-0.91	0.35	0.25		0.86
0.57		1.31	0.87	4.05	0.51	1.39	0.39	0.43	0.44	0.52	0.32	0.50			

Tabella A.9: Risultati delle simulazioni Monte Carlo realizzate con la seconda versione del metodo R-Average, per il primo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	s_0	t_0	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	t_{A_1}	t_{A_2}	t_{A_3}	t_{B_1}	t_{B_2}	t_{B_3}	R^2	
5	0.5	Mdn	2.96	-0.41	4.93	10.19	15.79	4.54	10.20	15.23	0.76	-0.73	0.60	-0.94	0.42	0.46	0.98
		Iqr	1.70	0.49	0.34	1.96	1.98	1.77	0.68	1.94	0.26	0.40	0.40	0.55	0.24	0.38	
	Mdn	3.25	-0.28	4.88	9.97	15.48	4.30	10.53	16.93	0.65	-0.67	0.72	-0.82	0.35	0.30	0.92	
	Iqr	1.94	0.52	0.71	3.32	3.31	2.69	1.91	4.94	0.59	0.81	0.80	0.90	0.66	0.75		
	Mdn	3.78	-0.09	4.59	8.99	18.04	4.45	11.17	18.42	0.00	0.00	0.30	-0.20	0.00	0.00	0.83	
	Iqr	2.50	0.53	1.47	1.97	5.81	2.36	2.16	3.30	0.65	0.56	1.09	0.70	0.41	0.32		
10	0.5	Mdn	2.58	-0.55	4.95	10.06	15.52	4.76	10.13	15.00	0.73	-0.70	0.60	-0.84	0.41	0.45	0.98
		Iqr	1.61	0.38	0.26	1.18	1.70	1.11	0.46	1.26	0.21	0.27	0.36	0.41	0.15	0.29	
	Mdn	3.20	-0.31	4.89	10.32	15.76	4.29	10.22	15.73	0.73	-0.75	0.64	-0.98	0.41	0.43	0.92	
	Iqr	1.86	0.49	0.50	2.78	2.27	2.45	1.00	3.14	0.34	0.56	0.50	0.62	0.33	0.49		
	Mdn	3.26	-0.27	4.90	9.71	15.46	4.41	10.70	17.45	0.63	-0.51	0.73	-0.77	0.30	0.23	0.83	
	Iqr	2.12	0.59	0.79	3.03	3.48	2.82	2.01	4.81	0.61	0.84	0.89	1.01	0.70	0.75		
15	0.5	Mdn	2.38	-0.59	4.95	10.05	15.36	4.86	10.11	14.94	0.72	-0.71	0.62	-0.80	0.40	0.44	0.98
		Iqr	1.37	0.32	0.22	0.85	1.46	0.86	0.38	1.03	0.17	0.19	0.32	0.34	0.13	0.25	
	Mdn	3.16	-0.37	4.89	10.27	15.74	4.38	10.19	15.40	0.73	-0.75	0.64	-0.96	0.42	0.43	0.91	
	Iqr	1.86	0.51	0.39	1.95	2.07	2.01	0.78	2.75	0.29	0.43	0.45	0.56	0.27	0.45		
	Mdn	3.27	-0.28	4.89	10.18	15.50	4.32	10.42	16.33	0.71	-0.71	0.71	-0.89	0.38	0.35	0.83	
	Iqr	1.84	0.52	0.60	3.14	2.52	2.42	1.55	4.71	0.40	0.67	0.64	0.75	0.47	0.68		

Tabella A.10: Risultati delle simulazioni Monte Carlo realizzate con la seconda versione del metodo R-Average, per il secondo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	Mdn	Iqr	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	t_{A_1}	t_{A_2}	t_{A_3}	t_{B_1}	t_{B_2}	t_{B_3}	R^2
5	0.5	Mdn		4.90	10.12	15.16	4.78	10.07	15.39	0.58	-0.77	0.59	-0.12	-0.11	-0.15	0.98
		Iqr		0.33	0.45	0.34	0.47	0.29	0.57	0.22	0.20	0.22	0.22	0.22	0.18	
	Mdn		4.65	10.23	15.51	4.20	10.13	16.55	0.58	-0.68	0.63	-0.19	-0.18	-0.18	-0.27	0.94
	Iqr		0.95	0.95	0.97	2.95	0.70	3.97	0.58	0.52	0.63	0.48	0.43	0.47		
	Mdn		4.20	10.27	16.13	2.85	10.27	17.80	0.28	0.00	0.26	-0.18	-0.15	-0.19	-0.39	0.86
	Iqr		3.80	1.16	4.08	3.15	1.22	3.40	0.67	0.88	0.60	0.39	0.36	0.39		
10	0.5	Mdn		4.91	10.08	15.12	4.80	10.04	15.31	0.57	-0.78	0.59	-0.12	-0.11	-0.14	0.98
		Iqr		0.21	0.30	0.22	0.31	0.20	0.36	0.14	0.14	0.14	0.14	0.13	0.14	
	Mdn		4.81	10.23	15.27	4.61	10.08	15.69	0.60	-0.76	0.66	-0.13	-0.14	-0.22	-0.37	0.93
	Iqr		0.57	0.68	0.58	1.08	0.48	1.75	0.40	0.32	0.36	0.35	0.32	0.37		
	Mdn		4.60	10.24	15.63	3.95	10.23	16.65	0.55	-0.66	0.52	-0.23	-0.19	-0.24	-0.43	0.86
	Iqr		1.03	0.96	1.17	3.23	0.79	3.82	0.66	0.62	0.67	0.41	0.42	0.43		
15	0.5	Mdn		4.93	10.06	15.10	4.81	10.03	15.27	0.58	-0.79	0.58	-0.12	-0.12	-0.14	0.98
		Iqr		0.16	0.24	0.16	0.23	0.17	0.26	0.10	0.11	0.11	0.09	0.10	0.10	
	Mdn		4.89	10.17	15.20	4.75	10.07	15.51	0.59	-0.77	0.61	-0.13	-0.11	-0.17	-0.28	0.93
	Iqr		0.40	0.53	0.39	0.63	0.36	0.81	0.26	0.24	0.26	0.26	0.21	0.28		
	Mdn		4.69	10.25	15.44	4.36	10.15	16.16	0.60	-0.70	0.65	-0.18	-0.16	-0.25	-0.45	0.86
	Iqr		0.79	0.88	0.83	2.42	0.65	3.48	0.51	0.45	0.54	0.43	0.39	0.45		

Tabella A.11: Risultati delle simulazioni Monte Carlo realizzate con la seconda versione del metodo R-Average, per il terzo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ	s_0	t_0	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	t_{A_1}	t_{A_2}	t_{A_3}	t_{B_1}	t_{B_2}	t_{B_3}	R^2	
5	0.5	Mdn	3.5	-0.23	4.84	10.28	15.79	4.64	10.41	15.79	0.74	-0.76	0.67	-0.12	-0.04	0.00	0.98
		Iqr	1.5	0.40	0.39	1.92	2.70	0.83	1.02	3.27	0.26	0.48	0.57	0.34	0.29	0.46	
	Mdn	4.08	0.00	4.51	8.83	18.98	4.49	10.51	16.33	0.43	-0.17	0.12	0.00	0.00	0.00	0.00	0.91
	Iqr	1.90	0.33	1.37	1.96	4.59	1.37	1.61	2.74	0.88	0.64	0.68	0.31	0.27	0.26		
	Mdn	4.33	0.00	3.89	8.38	19.93	4.47	10.42	16.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.81
	Iqr	1.86	0.00	1.80	1.71	1.39	1.77	1.82	2.25	0.00	0.00	0.00	0.00	0.00	0.00		
10	0.5	Mdn	3.12	-0.36	4.89	10.18	15.62	4.77	10.32	15.53	0.73	-0.73	0.68	-0.12	-0.03	-0.02	0.97
		Iqr	1.72	0.48	0.32	1.06	2.51	0.69	0.83	3.09	0.23	0.30	0.52	0.29	0.21	0.45	
	Mdn	3.80	-0.16	4.75	9.88	16.29	4.53	10.48	16.13	0.72	-0.65	0.57	-0.13	-0.01	0.00	0.00	0.90
	Iqr	1.57	0.34	0.62	3.15	3.77	1.15	1.43	3.59	0.43	0.91	0.69	0.38	0.40	0.50		
	Mdn	4.18	0.00	4.42	8.66	19.39	4.47	10.45	16.14	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.80
	Iqr	1.95	0.33	1.53	1.63	4.17	1.39	1.51	2.13	0.82	0.55	0.53	0.27	0.20	0.11		
15	0.5	Mdn	2.77	-0.48	4.91	10.08	15.42	4.87	10.24	15.18	0.72	-0.71	0.68	-0.08	-0.02	-0.02	0.97
		Iqr	1.76	0.51	0.26	0.81	1.98	0.56	0.75	2.69	0.21	0.22	0.42	0.28	0.17	0.40	
	Mdn	3.61	-0.17	4.83	10.34	15.93	4.59	10.45	15.92	0.75	-0.81	0.65	-0.10	-0.03	0.01	0.00	0.90
	Iqr	1.39	0.33	0.44	2.86	2.65	0.90	1.15	3.09	0.29	0.67	0.56	0.40	0.32	0.45		
	Mdn	3.88	-0.11	4.66	9.27	17.66	4.60	10.50	16.27	0.58	-0.42	0.37	-0.01	0.00	0.00	0.00	0.80
	Iqr	1.89	0.35	1.10	3.46	4.77	1.21	1.58	3.07	0.92	1.05	0.82	0.33	0.34	0.42		

Tabella A.12: Risultati delle simulazioni Monte Carlo realizzate con la seconda versione del metodo R-Average, per il quarto gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

$e^{t_0} < 0.1$			
Trials	σ_ϵ	Group 1	Group 3
5	0.5	790	783
	1.0	416	370
	1.5	227	218
10	0.5	872	887
	1.0	666	578
	1.5	338	328
15	0.5	923	963
	1.0	722	714
	1.5	227	218

Tabella A.13: Numero di modelli in cui il parametro w_0 è stimato inferiore a 0.1, per i due gruppi di parametri in cui lo stato iniziale è nullo, per la seconda versione del metodo R-Average.

Trials	σ_ϵ	s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	t_{A_1}	t_{A_2}	t_{A_3}	t_{B_1}	t_{B_2}	t_{B_3}	R^2
0.5	Mdn	5.01	10.01	14.99	4.99	9.99	15.02	0.57	-0.79	0.56	-0.81	0.24	0.24	0.98
	Iqr	0.24	0.27	0.26	0.27	0.23	0.28	0.14	0.18	0.15	0.17	0.17	0.17	
5	Mdn	4.99	10.06	15.02	5.06	9.97	15.09	0.51	-0.77	0.61	-0.74	0.22	0.18	0.94
	Iqr	0.46	0.54	0.56	0.60	0.46	0.59	0.33	0.36	0.36	0.39	0.36	0.36	
1.5	Mdn	4.91	10.11	15.06	5.12	9.92	15.08	0.32	-0.52	0.58	-0.51	0.00	0.00	0.87
	Iqr	0.82	0.84	0.91	0.94	0.75	0.84	0.68	1.01	0.80	0.83	0.60	0.62	
0.5	Mdn	5.01	10.01	15.01	5.01	9.99	15.02	0.56	-0.81	0.56	-0.81	0.25	0.25	0.98
	Iqr	0.17	0.19	0.18	0.19	0.16	0.19	0.09	0.13	0.09	0.12	0.10	0.10	
10	Mdn	5.01	10.02	14.97	5.00	9.99	15.08	0.56	-0.78	0.57	-0.80	0.23	0.21	0.93
	Iqr	0.34	0.39	0.37	0.41	0.33	0.39	0.18	0.25	0.20	0.25	0.23	0.22	
1.5	Mdn	4.99	10.08	14.97	5.04	9.95	15.05	0.52	-0.80	0.59	-0.75	0.21	0.20	0.86
	Iqr	0.54	0.59	0.57	0.66	0.51	0.58	0.39	0.42	0.40	0.42	0.38	0.38	
0.5	Mdn	5.01	9.99	15.00	5.01	10.00	15.01	0.56	-0.82	0.56	-0.82	0.25	0.25	0.98
	Iqr	0.13	0.16	0.15	0.17	0.12	0.15	0.06	0.10	0.07	0.10	0.07	0.07	
15	Mdn	5.02	9.99	14.99	5.01	9.99	15.04	0.56	-0.77	0.56	-0.80	0.23	0.21	0.93
	Iqr	0.28	0.33	0.30	0.33	0.25	0.31	0.16	0.20	0.16	0.20	0.20	0.19	
1.5	Mdn	4.99	10.03	14.98	5.00	9.97	15.05	0.54	-0.78	0.59	-0.78	0.22	0.19	0.86
	Iqr	0.43	0.49	0.46	0.49	0.39	0.46	0.26	0.34	0.26	0.31	0.27	0.27	

Tabella A.14: Risultati delle simulazioni Monte Carlo realizzate con la seconda versione del metodo R-Average, per il primo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. A differenza delle simulazioni riportate in tabella A.9, in questo caso le stime sono state effettuate escludendo dalla computazione i parametri di stato iniziale. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Trials	σ_ϵ		s_{A_1}	s_{A_2}	s_{A_3}	s_{B_1}	s_{B_2}	s_{B_3}	t_{A_1}	t_{A_2}	t_{A_3}	t_{B_1}	t_{B_2}	t_{B_3}	R^2
			5	10	15	5	10	15	0.58	-0.81	0.58	-0.12	-0.12	-0.12	
5	0.5	Mdn	4.99	9.99	15.01	5.03	10.01	14.95	0.58	-0.87	0.59	-0.10	-0.10	-0.10	0.98
		Iqr	0.26	0.28	0.24	0.28	0.24	0.28	0.11	0.20	0.11	0.08	0.08	0.08	
	1.0	Mdn	4.94	10.02	15.07	5.05	9.98	14.95	0.56	-0.84	0.56	-0.08	-0.07	-0.08	0.94
		Iqr	0.54	0.55	0.52	0.57	0.48	0.58	0.31	0.51	0.32	0.23	0.24	0.22	
	1.5	Mdn	4.73	10.04	15.27	5.09	10.02	14.92	0.37	-0.37	0.30	0.00	0.00	0.00	0.86
		Iqr	0.95	0.79	0.93	0.82	0.74	0.85	0.69	1.02	0.65	0.21	0.18	0.19	
10	0.5	Mdn	4.99	10.00	15.01	5.04	10.00	14.96	0.59	-0.87	0.59	-0.10	-0.10	-0.10	0.98
		Iqr	0.17	0.21	0.17	0.20	0.17	0.19	0.07	0.16	0.07	0.05	0.05	0.05	
	1.0	Mdn	4.99	10.01	15.01	5.05	9.99	14.96	0.59	-0.88	0.60	-0.09	-0.10	-0.10	0.93
		Iqr	0.35	0.42	0.36	0.40	0.33	0.40	0.19	0.32	0.20	0.12	0.12	0.11	
	1.5	Mdn	4.92	9.97	15.14	5.06	10.03	14.93	0.54	-0.82	0.52	-0.06	-0.05	-0.05	0.86
		Iqr	0.59	0.59	0.59	0.57	0.51	0.59	0.32	0.56	0.34	0.23	0.23	0.23	
0.5	Mdn	5.00	10.00	15.00	5.02	10.01	14.98	0.58	-0.85	0.58	-0.10	-0.10	-0.11	0.98	
	Iqr	0.15	0.17	0.14	0.16	0.15	0.16	0.05	0.12	0.05	0.05	0.05	0.05		
1.0	Mdn	5.01	10.00	15.01	5.05	10.00	14.97	0.59	-0.88	0.59	-0.10	-0.10	-0.10	0.93	
	Iqr	0.28	0.32	0.28	0.31	0.27	0.31	0.13	0.27	0.13	0.09	0.09	0.09		
1.5	Mdn	4.95	10.01	15.06	5.05	10.01	14.94	0.56	-0.83	0.56	-0.07	-0.08	-0.07	0.86	
	Iqr	0.45	0.44	0.46	0.48	0.42	0.55	0.22	0.44	0.25	0.16	0.17	0.17		

Tabella A.15: Risultati delle simulazioni Monte Carlo realizzate con la seconda versione del metodo R-Average, per il terzo gruppo di parametri (riportati in testa alla tabella) per tutte le condizioni di stima. A differenza delle simulazioni riportate in tabella A.11, in questo caso le stime sono state effettuate escludendo dalla computazione i parametri di stato iniziale. Ogni condizione è identificata dal numero di replicazioni e dalla componente d'errore $\mathcal{N}(0, \sigma^2)$ aggiunta ai valori R teorici. Per ogni condizione è indicata la mediana dei parametri stimati (Mdn) e lo scarto interquartile (Iqr). L'ultima colonna riporta l'adattamento del modello, espresso come media di R^2 per ogni simulazione eseguita.

Appendice B

Implementazione del metodo R-Average

Il metodo R-Average è stato implementato nella funzione *rav*, inserita all'interno della libreria *rAverage* (Vidotto et al., 2011) per il software statistico R. La libreria può essere scaricata dal sito web del progetto R e integrata all'interno del software. *rAverage* include una serie di funzioni programmate in linguaggio R integrato in diversi casi da istruzioni scritte in linguaggio C, che garantisce una maggior rapidità di calcolo.

La funzione *rav* implementa il metodo descritto nel cap. 2 e permette la manipolazione delle impostazioni descritte¹. A partire da una matrice di dati osservati, la funzione *rav* esegue la stima dei parametri e fornisce in output i modelli stimati. L'output è costituito da un oggetto definito attraverso lo standard *s4* per le classi R.

Per gli esempi d'uso e ulteriori approfondimenti si rimanda all'help delle funzioni descritte di seguito.

¹Attualmente, la versione backward della procedura Information Criteria non è implementata.

B.1 La funzione *rav* per la stima dei modelli averaging

Definizione

rav (data, subset = NULL, mean = FALSE, lev, range = NULL, start = c(s = NA, w = 1), lower = NULL, upper = NULL, I0 = TRUE, s.fixed = FALSE, w.fixed = FALSE, all = FALSE, IC.break = FALSE, IC.diff = c(2, 2), Dt = 0.1, method = L-BFGS-B, control = list(), title = NULL, names = NULL, verbose = FALSE)

Argomenti

- data** Oggetto di tipo matrice, data frame o vettore contenente i dati. Ogni colonna corrisponde a un disegno del piano fattoriale (in ordine: disegni a una via, a due vie, ecc.). Le colonne devono essere ordinate combinando ogni livello del primo fattore considerato, con tutti i livelli dei fattori successivi. La prima colonna può contenere un vettore che descrive un codice identificativo del soggetto.
- subset** Vettore di tipo numerico, carattere o fattore, che individua un sottoinsieme di soggetti su cui svolgere l'analisi.
- mean** Valore logico che specifica se eseguire l'analisi sulle medie calcolate per ogni disegno invece che sui dati grezzi.
- lev** Vettore contenente il numero di livelli di ogni fattore.
- range** Vettore contenente la gamma entro cui possono variare le risposte osservate, ovvero i limiti minimo e massimo della scala di risposta. Se NULL, saranno utilizzati il valore minimo e massimo delle risposte osservate come limiti della scala. Il valore assunto da questo argomento determina la gamma di variazione dei parametri s quando viene utilizzata la routine di minimizzazione L-BFGS-B.

- start** Vettore contenente gli starting values il valore di scala s e per i pesi w . Per default, lo start per s è costituito dal valore di centro scala e per w il valore 1.
- lower** Vettore contenente il valore minimo che i parametri s e w possono assumere durante il processo di ottimizzazione quando viene utilizzata la routine L-BFGS-B. Se NULL, per s sarà utilizzato il limite inferiore della scala e per w il valore $\exp(-4)$.
- upper** Vettore contenente il valore massimo che i parametri s e w possono assumere durante il processo di ottimizzazione quando viene utilizzata la routine L-BFGS-B. Se NULL, per s sarà utilizzato il limite superiore della scala e per w il valore $\exp(4)$.
- IO** Valore logico che indica che i parametri di stato iniziale s_0 e w_0 devono essere considerati o meno. Se TRUE, i due parametri saranno stimati, altrimenti saranno forzati a essere zero.
- s.fixed** Valore logico che indica se i valori di scala devono essere stimati durante il processo di ottimizzazione. Se TRUE, una maschera per l'inserimento dei loro valori comparirà sullo schermo, e alcuni o tutti gli s potranno essere fissati a un determinato valore. In alternativa, può essere specificato un oggetto di tipo *list* che ne contiene i valori fissi. Se l'argomento *s.fixed* è pari a due, le medie dei sotto-disegni a una via saranno utilizzati come parametri s .
- w.fixed** Valore logico che indica se i valori dei pesi devono essere stimati durante il processo di ottimizzazione. Se TRUE, una maschera per l'inserimento dei loro valori comparirà sullo schermo, e alcuni o tutti i w potranno essere fissati a un determinato valore. In alternativa, può essere specificato un oggetto di tipo *list* che ne contiene i valori fissi.
- all** Valore logico che indica se utilizzare la versione completa della procedura Information Criteria. Se FALSE, verrà utilizzata la versione ridotta.

- IC.break** Valore logico che indica se avviare la stima tramite la procedura Information Criteria. Se FALSE, il modello Information Criteria non sarà stimato.
- IC.diff** Vettore contenente i valori soglia ΔBIC e ΔAIC entro cui modelli diversi devono essere considerati equivalenti.
- Dt** Valore numerico che specifica Δ_t , il valore soglia entro cui diversi parametri t devono essere considerati uguali.
- method** Routine di minimizzazione da utilizzare. Le opzioni sono: “L-BFGS-B”, “BFGS”, “Nelder-Mead”, “SANN” e “CG”. Si rimanda alla documentazione della funzione *optim* per ulteriori dettagli (R Development Core Team, 2011).
- control** Una lista di controlli da passare alla funzione *optim* durante il processo di ottimizzazione. Si rimanda alla documentazione della funzione *optim* della libreria *stats* per ulteriori dettagli.
- title** Stringa di carattere che specifica un titolo da attribuire all’output.
- names** Vettore di stringhe di carattere che specificano i nomi dei fattori, nell’ordine in cui sono stati inseriti nell’argomento *data*.
- verbose** Valore logico. Se TRUE, la funzione stamperà alcune informazioni riassuntive per ogni passo della procedura Information Criteria.

B.2 Principali funzioni di supporto all’analisi

- averaging** A partire da un gruppo di parametri genera le risposte osservate R secondo la funzione *averaging*.
- outlier.remove** Individua e rimuove i valori anomali.
- rav.fitted** Estrae i valori predetti dal miglior modello stimato dalla funzione *rav*.
- rav.indexes** Calcola gli indici di adattamento di un modello *averaging*.
- rav.param** Estrae i parametri del miglior modello stimato con la funzione *rav*.
- rav.resid** Estrae i residui del miglior modello stimato con la funzione *rav*.

Appendice C

Credenze di fiducia: item proposti

C.1 Ambito individuale

1. Tua madre si sta risposando dopo un divorzio. Questo comporta che il suo nuovo marito verrà a vivere nella vostra casa. Ti stai chiedendo come funzionerà la vostra convivenza. Quanta fiducia riponi nel nuovo compagno di tua madre che presenta le seguenti caratteristiche?
2. Tuo figlio adolescente ti chiede di comprargli il motorino. Poiché sa che tu lo consideri pericoloso, ti assicura che rispetterà ogni regola di prudenza. Quanta fiducia riponi in tuo figlio sapendo che presenta le seguenti caratteristiche?
3. Hai fatto una confidenza a un amico. Ti auguri che la confidenza rimanga tale. Quanta fiducia riponi nel tuo amico sapendo che presenta le seguenti caratteristiche?
4. Stai partendo per le vacanze. Hai curato con tanta attenzione le piante nel tuo appartamento. Stai pensando che il tuo vicino di casa sarebbe disponibile a innaffiarle. Devi però lasciargli le chiavi di casa. Quanta fiducia riponi nel tuo vicino sapendo che presenta le seguenti caratteristiche?

5. Cerchi qualcuno con cui condividere le spese del tuo appartamento. Alcune persone rispondono al tuo annuncio. Quanta fiducia riponi in una persona che presenta le seguenti caratteristiche?
6. Stai cercando una donna che venga a fare alcune ore settimanali di pulizia nella tua casa. Ti si presenta una signora che non conosci personalmente. Quanta fiducia riponi in una signora che presenta le seguenti caratteristiche?

C.2 Ambito istituzionale

1. Chiedi informazioni a uno sportello in stazione su come effettuare un lungo viaggio che richiede alcuni cambi di linea. Il viaggio comprenderà anche un'intera notte per cui dovrai utilizzare una carrozza letto o una cuccetta. È la prima volta che fai questo viaggio. Hai timore che l'addetto alla biglietteria possa sbagliare le coincidenze. Quanta fiducia riporresti in un incaricato che sai essere descritto dalle seguenti caratteristiche?
2. Ti rechi negli uffici della Regione perché vuoi conoscere se le previste modifiche al tuo impianto di riscaldamento potranno godere di un contributo pubblico. Quanta fiducia riponi in un dipendente della Regione che presenta le seguenti caratteristiche?
3. Sono vicine le elezioni per la nomina del sindaco del tuo comune. Quanta fiducia riporresti in un candidato che presenta le seguenti caratteristiche?
4. Ti sei appena diplomato, uno dei tuoi insegnanti ti suggerisce il tipo di università per il quale ti vede più portato. Quanta fiducia riponi nel tuo insegnante sapendo che egli presenta le seguenti caratteristiche?
5. Hai un persistente dolore a un ginocchio. Lo specialista che ti visita ritiene necessaria un'operazione chirurgica. Quanta fiducia riponi in un medico che presenta le seguenti caratteristiche?

6. Nella tua chiesa viene lanciata una raccolta di fondi per l'invio di medicinali in un paese straniero colpito dalla carestia. Quanta fiducia riponi nel sacerdote promotore sapendo che presenta le seguenti caratteristiche?

C.3 Ambito organizzativo

1. Il settore nel quale opera la tua azienda è in contrazione. Alcune aziende come la tua hanno già chiuso. Ti interroghi sul tuo futuro. L'imprenditore ti assicura che l'azienda supererà il momento di crisi. Quanta fiducia riponi nell'imprenditore sapendo che presenta le seguenti caratteristiche?
2. Ti è chiesto di lavorare a un progetto assieme a una persona con la quale non hai mai lavorato prima. Quanta fiducia riponi in un collaboratore che presenta le seguenti caratteristiche?
3. Il tuo capo ti dice che se accetterai di lavorare in trasferta e fare straordinari per un certo tempo (2-3 anni) avrai un'importante promozione. Quanta fiducia riponi in un capo che presenta le seguenti caratteristiche?
4. Vai a ritirare il computer che avevi portato a riparare. Ti viene detto che sono stati sostituiti alcuni elementi costosi del PC. Ti chiedi se tali sostituzioni siano state effettivamente effettuate/necessarie. Quanta fiducia riponi in un tecnico che presenta le seguenti caratteristiche?
5. Sei incerto tra due impianti stereo di uguali caratteristiche ma di marche diverse. Il commesso ti raccomanda uno dei due perché afferma abbia una resa sonora migliore. Quanta fiducia riponi in un commesso che presenta le seguenti caratteristiche?
6. Hai accettato un appuntamento per la prima volta con il rappresentante di una ditta che ti propone dei prodotti ai quali sei interessato. Quanta fiducia riponi in un venditore che presenta le seguenti caratteristiche?

Bibliografia

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Almpanidis, G., & Kotropoulos, C. (2008). Phonemic segmentation using the generalised Gamma distribution and small sample Bayesian information criterion. *Speech Communication*, 50(1), 38-55.
- Anderson, N. H. (1959). Test of a model for opinion change. *Journal of Abnormal and Social Psychology*, 59(3), 371-381.
- Anderson, N. H. (1962). Application of an additive model to impression formation. *Science, New Series*, 138(3542), 817-818.
- Anderson, N. H. (1965). Averaging versus adding as a stimulus combination rule in impression formation. *Journal of Experimental Psychology*, 70(4), 394-400.
- Anderson, N. H. (1967). Averaging model analysis of set-size effect in impression formation. *Journal of Experimental Psychology*, 75(2), 158-165.
- Anderson, N. H. (1970). Functional measurement and psychophysical judgement. *Psychological Review*, 77(3), 153-170.
- Anderson, N. H. (1972). Looking for configurality in clinical judgment. *Psychological Bulletin*, 78(2), 93-102.
- Anderson, N. H. (1973). Comments on the articles of Hodges and of Schönemann, Cafferty and Rotton. *Psychological Review*, 80(1), 88-92.

- Anderson, N. H. (1977). Note on functional measurement and data analysis. *Perception & Psychophysics*, *21*(3), 201-215.
- Anderson, N. H. (1981). *Foundation of information integration theory*. New York: Academic Press.
- Anderson, N. H. (1982). *Methods of information integration theory*. New York: Academic Press.
- Anderson, N. H. (1991). *Contributions to information integration theory*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, N. H., & Butzin, C. A. (1974). Performance = motivation \times ability: an integration-theoretical analysis. *Journal of Personality and Social Psychology*, *30*(5), 598-604.
- Anderson, N. H., & Shanteau, J. C. (1970). Information integration in risky decision making. *Journal of Experimental Psychology*, *84*(3), 441-451.
- Bellizzi, J. A., & Crowley, A. E. (1983). The effects of color in store design. *Journal of retailing*, *59*(1), 21-45.
- Bogdan, M., Ghosh, J. K., & Doerge, R. W. (2004). Modifying the Schwarz Bayesian Information Criterion to Locate Multiple Interacting Quantitative Trait Loci. *Genetics*, *167*(2), 989-999.
- Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. Willey Oxford, England: John Wiley & Sons.
- Bromiley, P., & Cummings, L. L. (1995). Transactions costs in organizations with trust. In R. Bies, R. Lewicki, & B. Sheppard (Eds.), *Research on negotiations in organizations* (pp. 219-247). Greenwich: JAI Press.
- Broyden, C. G. (1970). The Convergence of a Class of Double-rank Minimization Algorithms. *IMA Journal of Applied Mathematics*, *6*(1), 76-90.

- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 3(2), 261-304.
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *Journal Scientific Computing*, 16(6), 1190-1208.
- Carmone, F. J., & Green, P. E. (1981). Model misspecification in multiattribute parameter estimation. *Journal of Marketing Research*, 18(1), 87-93.
- Carroll, J. D., & Green, P. E. (1995). Psychometric methods in marketing research: part I, conjoint analysis. *Journal of Marketing research*, 32(4), 385-391.
- Chandler, J. P. (1969). STEPIT - finds Local Minima of a Smooth Function of Several Parameters. *Behavioral Science*, 14(1), 81-82.
- Cummings, L. L., & Bromiley, P. (1996). The Organizational Trust Inventory (OTI): Development and validation. In R. Kramer, & T. Tyler (Eds.), *Trust in organizations* (pp. 302-330). Thousand Oaks, CA: Sage.
- Dougherty, M. R. P., & Shanteau, J. (1999). Averaging expectancies and perceptual experiences in the assessment of quality. *Acta Psychologica*, 101(1), 49-67.
- Fletcher, R. (1970). A New Approach to Variable Metric Algorithms. *The Computer Journal*, 13(3), 317-322.
- Friedman, M., & Savage, L. J. (1952). The expected-utility hypothesis and the measurability of utility. *Journal of Political Economy*, 60(6), 463-474.
- Goldfarb, D. (1970). A Family of Variable Metric Updates Derived by Variational Means. *Mathematics of Computation*, 24(109), 23-26.

- Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8(3), 355-363.
- Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: issues and outlook. *Journal of Consumer Research*, 5(2), 103-123.
- Hurvich, C. M., & Tsai, C. L. (1989). Regression and Time Series Model Selection in Small Samples. *Biometrika*, 76(2), 297-307.
- Janiszewski, C., & Cunha, M. C. (2004). The influence of price discount framing on the evaluation of a product bundle. *Journal of Consumer Research*, 30(4), 534-546.
- Jo, M. S., Nakamoto, K., & Nelson, J. E. (2003). The shielding effects of brand image against lower quality countries-of-origin in global manufacturing. *Journal of Business Research*, 56(8), 637-646.
- Johar, G. V., Jedidi, K. J., & Jacoby, J. (1997). A varying-parameter averaging model of on-line brand evaluations. *Journal of Consumer Research*, 24(2), 232-247.
- Krantz, D. H., & Tversky, A. (1978). Conjoint measurement analysis of composition rules in psychology. *Psychological Review*, 78(2), 151-169.
- Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. *Sociological Methods & Research*, 33(2), 188-229.
- Lynch, J. G. Jr. (1985). Uniqueness issues in the decompositional modeling of multiattribute overall evaluations: an information integration perspective. *Journal of Marketing Research*, 22(1), 1-19.
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *Journal of Mathematical Psychology*, 1(1), 1-27.

- Massidda, D., Noventa, S., & Vidotto, G. (2009). Un modello a media ponderata per l'analisi dei giudizi multi-attributo. *Quaderni di Management*, 39.
- McKnight, D. H., & Chervany, N. L. (2000). What is Trust? A Conceptual Analysis and an Interdisciplinary Model. *AMCIS 2000 Proceedings*, Paper 382. Retrieved from <http://aisel.aisnet.org/amcis2000/382>.
- McKnight, D. H., Cummings, L. L., & Chervany, N. L. (1998). Initial trust formation in new organizational relationships. *Academy Of Management Review*, 23(3), 473-490.
- Meyvis, T., & Janiszewski, C. (2002). Consumers' beliefs about product benefits: the effect of obviously irrelevant product. *Journal of Consumer Research*, 28(4), 618-635.
- Moskowitz, H. R., & Silcher, M. (2006). The application of conjoint analysis and their possible uses in sensometrics. *Food Quality and Preference*, 17(3-4), 145-165.
- Myers, J. H., & Alpert, M. I. (1977). Semantic confusion in attitude research: salience vs. importance vs. determinance. *Advances in Consumer Research*, 4(1), 106-10.
- Norman, K. L. (1976). A solution for weights and scale values in functional measurement. *Psychological Review*, 83(1), 80-84.
- Noventa, S., Massidda, D., & Vidotto, G. (2010). Initial conditions in the averaging cognitive model. *Psicológica*, 31(3), 477-489.
- Noventa, S., Massidda, D., & Vidotto, G. (in revisione). Is there a need for more than three models? *Psicológica*.
- Pearse, N. (2011). The Granularity of Scale Response Categories: The use of a 21-Point Scale. In Ashwin, M. (Ed.), *The Proceedings of the 10th European Conference on Research Methodology for Business and Management Studies* (pp. 420-429). Normandy Business School, Caen, France.

- R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-163.
- Rotter, J. B. (1967). A new scale for the measurement of interpersonal trust. *Journal of personality*, 35(4), 651-665.
- Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
- Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation*, 24(111), 647-656.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461-464.
- Singh, R. (1991). Two problems in cognitive algebra: imputations and averaging versus multiplying. In: N. H. Anderson (Ed.), *Contributions to Information Integration Theory* (Vol. II: Social, pp. 143-180). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Skowronski, J. J., & Carlston, D. E. (1987). Social Judgment and Social Memory: The Role of Cue Diagnosticity in Negativity, Positivity, and Extremity Biases. *Journal of Personality and Social Psychology*, 52(4), 689-699.
- Struyf, A., Hubert, M., & Rousseeuw, P. J. (1997). Integrating robust clustering techniques in S-PLUS. *Computational Statistics & Data Analysis*, 26(1), 17-37.
- Sugiura, N. (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communication in Statistics - Theory and Methods*, 7(1), 13-26.
- Tremblay, M., & Wallach, D. (2004). Comparison of parameter estimation methods for crop models. *Agronomie*, 24(6-7), 351-365.

- Troutman, C. M., & Shanteau, J. (1976). Do consumers evaluate products by adding or averaging attribute information? *Journal of consumer research*, 3(2), 101-106.
- Van Ittersum, K., Pennings, J. M. E., Wansink, B., & van Trijp, H. C. M. (2007). The validity of attribute-importance measurement: A review. *Journal of Business Research*, 60(11), 1177-1190.
- Vidotto, G. (2011). Note on Differential Weight Averaging Models in Functional Measurement. *Quality & Quantity*, on line first DOI 10.1007/s11135-011-9567-1.
- Vidotto, G., Massidda, D., & Noventa, S. (2010). Averaging models: parameters estimation with the R-Average procedure. *Psicologica*, 31(3), 461-475.
- Vidotto, G., Noventa, S., Massidda, D., & Vicentini, M. (2011). *rAverage: Parameter estimation for the Averaging model of Information Integration Theory* [Computer program]. Retrieved from <http://www.r-project.org>.
- Vidotto, G., & Vicentini, M. (2007). A general method for parameter estimation of averaging models. *Teorie & modelli*, 12(1-2), 211-221.
- Vidotto, G., Vicentini, M., Argentero, P., & Bromiley, P. (2008). Assessment of organizational trust: Italian adaptation and factorial validity of the organizational trust inventory. *Social Indicators Research*, 88(3), 563-575.
- Wang, M., & Yang, J. (1998). A multi-criterion experimental comparison of three multi-attribute weight measurement methods. *Journal of Multicriteria Decision Analysis*, 7(6), 340-350.
- Zalinski, J. (1984). Parameter estimation: An estimated variance-covariance matrix for averaging model parameters in information integration theory. *Behavior Research Methods, Instruments, & Computers*, 16(6), 557-558.

- Zalinski, J. (1987). *Parameter estimation for the averaging model of information integration theory*. Doctoral dissertation, University of California, San Diego, La Jolla, CA.
- Zalinski, J., & Anderson, N. H. (1986). *Average: a user-friendly fortran-77 program for parameter estimation for the averaging model of information integration theory* [Computer program]. San Diego, CA.
- Zalinski, J., & Anderson, N. H. (1989). Measurement of importance in multi-attribute models. In: J. B. Sidowski (Ed.), *Conditioning, cognition, and methodology. Contemporary issues in experimental psychology* (pp. 177-215). Lanham, MD: University press of America.
- Zalinski, J., & Anderson, N. H. (1991). Parameter estimation for averaging theory. In N. H. Anderson (Ed.), *Contributions to Information Integration Theory* (Vol. I: Cognition, pp. 353-394). Hillsdale, NJ: Lawrence Erlbaum Associates.