

**UNIVERSITA' DEGLI STUDI DI PADOVA**

**FACOLTA' DI SCIENZE STATISTICHE  
CORSO IN SCIENZE STATISTICHE ED ECONOMICHE**

**TESI DI LAUREA**

**LA METODOLOGIA NPC RANKING NELLA CONJOINT ANALYSIS E  
APPLICAZIONI AL PRODOTTO BANCARIO CONTO GIOVANI**

**Relatore: Ch.mo Prof. Fortunato Pesarin**

**Correlatore: Ch.mo Prof. Luigi Salmaso**

**Laureando: Tommaso Fabris**

**ANNO ACCADEMICO 2002-2003**



*Ai miei genitori,  
per tutto il loro affetto e sostegno*

*Un ringraziamento sentito ai professori Fortunato Pesarin e Luigi Salmaso per la loro  
pazienza e disponibilità.*

*Sento di dovere un ringraziamento speciale a tutta la mia famiglia e a tutte le persone  
che mi hanno voluto bene, in particolare a:*

*Silvia, per il tuo amore che mi rende ogni giorno felice e migliore*

*Marco, per la tua eterna amicizia*

*Alessandro e Pietro, per la vostra sincera amicizia*

*Samuele, perché su di te posso sempre contare*

*Francesco e Paolo, per la vostra simpatia e amicizia*

*Marco e Alessia, per il vostro grande affetto*

*Rino e Antonella, per il vostro sincero affetto*

*Enrico, per la lealtà e perché posso dire di aver trovato un vero amico*

*Moreno, per il tuo prezioso aiuto*



# Indice

<b>PREFAZIONE</b>		<b>9</b>
<b>CAPITOLO 1</b>	<b>ANALISI DELLA LETTERATURA SULLA CONJOINT ANALYSIS</b>	<b>13</b>
1.1	Introduzione	13
1.2	Sviluppo della Conjoint Analysis negli anni '80	16
1.3	Sviluppo della Conjoint Analysis negli anni '90	20
1.4	Applicazioni	20
1.5	Prospettive future	25
1.6	Riflessioni e prospettive	27
1.7	I modelli della Conjoint Analysis	32
1.8	Presentazione degli stimoli	35
<b>Capitolo 2</b>	<b>CONJOINT ANALYSIS: FULL PROFILE</b>	<b>37</b>
2.1	Introduzione	37
2.2	I concetti fondamentali e le fasi di un esperimento di Conjoint Analysis	39
2.3	Specificazione e stima del modello di utilità individuale	43
2.4	Analisi grafica dell'influenza degli effetti dei fattori sulla preferenza dei rispondenti	47
2.5	Verifica di Ipotesi sugli effetti degli attributi mediante Anova	49
2.6	L'uso di piani fattoriali nella Conjoint Analysis	52
2.7	Analisi bibliografica	55
2.8	Alcuni problemi sull'aggregazione dei rispondenti nella Conjoint Analysis	58

<b>Capitolo 3</b>	<b>APPLICAZIONI DELLA CONJOINT ANALYSIS NEL SETTORE BANCARIO</b>	<b>61</b>
3.1	Introduzione	61
3.2	Gli investimenti finanziari	62
3.3	Le banche	74
<b>CAPITOLO 4</b>	<b>METODI DI AGGREGAZIONE DEI RISPONDENTI E PROCEDURE NON PARAMETRICHE PER LA VERIFICA D'IPOTESI DI "COERENZA"</b>	<b>83</b>
4.1	Introduzione	83
4.2	Presentazione del problema	83
4.3	Condizioni di applicabilità del metodo	84
4.4	Metodologia della procedura	85
	4.4.1 Costruzione delle classifiche parziali	85
	4.4.2 Costruzione della classifica finale	86
4.5	Confronto tra graduatorie	87
	4.5.1 L'indice di Spearman $I_S$ ed il test TS	88
	4.5.2 L'indicatore di permutazione sulle graduatorie $I_{PG}$ ed il test TPG	89
4.6	I Passi dell'algoritmo	93
4.7	Implementazione in Sas	96
	4.7.1 Introduzione	96
	4.7.2 Graduatoria ottenuta con l'NPC ranking	98
	4.7.3 Graduatoria ottenuta con la media	101
	4.7.4 Indice di $I_S$ per l'NPC ranking	102
	4.7.5 Indice di $I_S$ per la media	104
	4.7.6 Indicatore $I_{PG}$ per l'NPC ranking	106
	4.7.7 L'indicatore $I_{PG}$ per la media	111

	4.7.8	Confronto dei risultati	116
4.8		Clusters Analysis: implementazione dell'algoritmo di Zani per le classificazioni "sfocate"	118
	4.8.1	Implementazione in Sas	119
<b>Capitolo 5</b>		<b>STUDIO COMPARATIVO DI SIMULAZIONE: NPC RANKING VS MEDIA</b>	<b>127</b>
	5.1	Introduzione	127
	5.2	Cardinalità dello spazio delle permutazioni	128
	5.3	Descrizione dello studio di simulazione	130
	5.4	Studio di simulazioni	134
	5.4.1	Comportamento con errori di tipo Normale	135
	5.4.2	Comportamento con errori di tipo Gamma	138
	5.4.3	Comportamento con errori di tipo Uniforme	142
	5.4.4	Comportamento con errori di tipo Binomiale	145
	5.5	Studio di simulazione per i test TPG e TS	148
	5.6	Potenza dei test sotto H1	150
<b>Capitolo 6</b>		<b>APPLICAZIONI ALLO STUDIO PILOTA DI CONJOINT ANALYSIS SUL PRODOTTO BANCARIO: CONTO GIOVANI</b>	<b>153</b>
	6.1	Indagine preliminare	153
	6.2	Svolgimento dell'indagine	153
	6.3	Modalità d'intervista	157
	6.4	Scelta ed analisi del campione	157
<b>APPENDICE</b>		<b>SAS MACRO LANGUAGE</b>	<b>163</b>
<b>BIBLIGRAFIA</b>			<b>185</b>



## Prefazione

Questa tesi ha come obiettivo principale quello di proporre l'utilizzo, nell'ambito della Conjoint Analysis (C.A), della combinazione non parametrica di graduatorie (NPC ranking) e saggiarne la validità attraverso confronti e simulazioni.

La C.A. è nata negli anni '70 e da allora le sue notevoli caratteristiche di flessibilità e adattabilità hanno fatto sì che essa trovasse numerose applicazioni e fosse sviluppata in molti settori, in particolare nel marketing. Nella tesi sarà presa in considerazione prevalentemente un'applicazione nel marketing bancario.

Questa tecnica è in continua evoluzione e si avvale di svariate procedure per la raccolta dei dati, la più diffusa è quella chiamata FULL PROFILE che tiene conto di tutti gli attributi di un prodotto.

La C.A. consente di comunicare le preferenze espresse dai consumatori (unità statistiche) per un determinato prodotto, attraverso la specificazione di alternative multiattributive (cartellini) dello stesso e di aggregarle in un'unica classifica finale utilizzando la media aritmetica (semplice o ponderata). Una lunga ricerca bibliografica e consultazioni con specialisti del campo hanno confermato che, la media, è l'unico modo oggi usato in C.A. per costruire una classifica finale.

Un'alternativa alla media può essere rappresentata dall'NPC ranking che è un metodo applicabile a problemi in cui si vogliono classificare delle variabili, in base ai giudizi di più persone intervistate. Questo metodo è utilizzato per combinare una pluralità di classifiche parziali in un'unica finale.

Si sono introdotti due indici di confronto: l'indice di Spearman  $I_S$  e l'indicatore di Permutazione delle Graduatorie  $I_{PG}$ , per comparare questo nuovo metodo con la media. Oltre all' $I_S$  e all' $I_{PG}$  si sono studiati due test: un particolare test di Spearman TS ed il test di permutazione sulle graduatorie TPG.

Successivamente si sono implementati in SAS macro language i passi degli algoritmi per il calcolo dell'NPC ranking e della media, dell' $I_S$  e dell' $I_{PG}$ .

Questa parte del lavoro relativo alla programmazione è particolarmente importante perché permette di applicare e confrontare in poco tempo e agevolmente l'NPC ranking e la media.

Si sono poi effettuate delle simulazioni generando dei dati sommando ad una graduatoria di riferimento errori originati dalle seguenti distribuzioni discrete e continue: Normale, Gamma, Uniforme e Binomiale.

I risultati delle simulazioni mettono in evidenza che l’NPC Ranking risulta più affidabile della media quando si ipotizzano giudizi dei rispondenti lontani dalla graduatoria di riferimento, mentre nel caso contrario media ed NPC Ranking risultano avere una affidabilità simile. Tutti questi esperimenti sono stati analizzati utilizzando il programma SAS, che permette di lavorare con grandi quantità di dati e di conseguenza disporre di risultati estremamente attendibili. Il caso simulato presumeva otto “cartellini” e mille gruppi, di cinque persone ognuno. Inoltre si sono calcolati gli  $I_S$  e gli  $I_{PG}$  di tutte le possibili permutazioni della graduatoria presa come riferimento nelle simulazioni (cioè  $8!=40320$   $I_S$  ed altrettanti  $I_{PG}$ ).

Si è poi calcolata la potenza dei test TS e TPG sotto l’ipotesi alternativa, generando i dati da una classifica diversa da quella di riferimento, ma solo leggermente modificata. I due test si sono rivelati potenti in quanto segnalano già ad  $\alpha$  nominali bassi che la classifica utilizzata per le simulazioni, sotto l’ipotesi alternativa, è diversa da quella di riferimento.

Infine si riporta lo studio di un caso reale riguardante il settore bancario. In particolare si è svolta un’analisi preliminare per alcune banche venete allo scopo di capire se il settore bancario potesse essere realmente interessato a sviluppare altri eventuali progetti sfruttando le potenzialità della C.A.. Questa indagine preliminare riguarda un tipico e diffuso servizio bancario: il ‘conto giovani’, e si propone di individuarne le caratteristiche principali, utilizzando l’NPC ranking per aggregare le risposte degli intervistati (di età inferiore a 29 anni).

**Gli argomenti qui sopra trattati sono suddivisi in sei capitoli:**

- i primi tre analizzano la C.A. sotto gli aspetti teorici, bibliografici e storici.
- il quarto è quello che introduce l’NPC ranking, gli indici  $I_S$  e  $I_{PG}$  e i due test TS e TPG e la loro relativa implementazione in SAS.
- il quinto contiene l’analisi e i risultati delle simulazioni.
- il sesto descrive il caso reale sul “conto giovani”.

I risultati principali ottenuti in questo lavoro sono che:

- In letteratura la media è ritenuta l'unica soluzione per aggregare i giudizi di preferenza nella Conjoint Analysis.
- L'NPC ranking è una valida alternativa alla media.
- Si è presentata una nuova strategia per il confronto di graduatorie ( $I_{PG}$ ), che utilizza una procedura di verifica di ipotesi per misurare la coerenza tra le graduatorie.
- I due test TS e TPG si sono rivelati potenti in quanto segnalano già ad  $\alpha$  nominali bassi che la classifica utilizzata per le simulazioni, sotto l'ipotesi alternativa, è diversa da quella di riferimento.
- Si sono implementati in SAS macro language i passi degli algoritmi per il calcolo dell'NPC ranking e della media, dell' $I_S$  e dell' $I_{PG}$  e dei test TS e TPG. Inoltre si è svolto uno studio comparativo di simulazione.



## *Capitolo 1*

# ***ANALISI DELLA LETTERATURA SULLA CONJOINT ANALYSIS***

### **1.1 Introduzione**

La Conjoint Analysis (C.A.) si è sviluppata grazie ai numerosi lavori inerenti alla raccolta di dati e al “modelling” svolti da “marketing researchers”. Il 1970 ha portato una ricca collezione di strumenti e tecniche. La fine degli anni '60 e l'inizio degli anni '70 sono caratterizzati da un forte interesse per i metodi di ricerca di mercato da parte delle scienze comportamentali, prima di tutto psychometrics e poi mathematical psychology. In questo periodo si sviluppano tre particolari tecniche di ricerca di mercato:

- CLUSTER ANALYSIS
- MULTIMENSIONAL SCALING [MDS], spesso chiamata PERCEPTUAL AND PREFERENCE MAPPING
- CONJOINT ANALYSIS

Il CLUSTER ANALITIC METHODS, ha trovato applicazione, quasi immediatamente, nel “MARKET SEGMENTATION” [Green, Frank e Robinson, 1967]. I ricercatori assumono di conoscere le caratteristiche dei segmenti identificati a priori nello studio, e, grazie alla CLUSTER ANALYSIS, sono in grado di valutare l'esattezza degli assunti iniziali. Potevano, sulla base della CLUSTER BASED SEGMENTATION, cercare i bisogni e i benefici, identificare le preferenze di marchio e analizzare alternative di

possibili ai problemi. L'idea della segmentazione si identificava nell'isolamento di gruppi di consumatori con uguali bisogni, attitudini, trade-offs e benefici.

MDS, MULTIMENSIONAL SCALING, ha ricevuto considerevole attenzione nel 1970 da parte dei ricercatori di marketing, che sostenevano l'importanza della costruzione di un progetto per rappresentare in due o più modi profili figurativi o tavole numeriche come punto o punti-vettori geometrici in due dimensioni.

Era nata anche la possibilità di trarre vantaggio dalla combinazione delle MDS con l'HIERARCHICAL CLUSTERING METHODS per aumentare le configurazioni bidimensionali con il CLUSTER BASED REPRESENTATIONS, basato su soluzioni in alta dimensionalità [Carroll e Green, 1997].

I ricercatori che volevano usare l'MDS per sviluppare nuovi prodotti o risollevere prodotti stagnanti, si trovarono davanti a due grossi problemi:

1. tradurre il punteggio del brand su dimensioni che potevano essere manipolate
2. avere livelli e attributi manipolabili in percentuale o preferenza

Le precedenti questioni illustrarono il problema generale del Reverse Engineering, cioè la traduzione inversa da percezioni di attributi a caratteristiche fisiche e chimiche del prodotto, che non sono tipicamente uno a uno [Kaul e Rao, 1994].

E' un piccolo prodigio che la C.A., con la sua enfasi sulla ricerca di specifici livelli di attributi, provvede a basi per le preferenze relative a scelte dei rispondenti su specifici argomenti, che includono attributi fisici ("verbal descriptions") di prodotto e di servizio. I primordiali precursori che scrissero documenti inerti alla C.A. furono Luce and Taker, ed il primo articolo intitolato "Conjoint Measurement" venne pubblicato nel 1964 sul "Journal of Mathematical Psychology". Gli autori si focalizzarono su approcci assiomatici per misure fondamentali: l'idea era di ottenere dati su una scala metrica ordinata attraverso un "rank-order response date" e un profilo fattoriale.

Non sorprende, che il primo algoritmo di C.A., chiamato MONANOVA, disegnato da Kruskal [1965] e programmato da Kruskal e Carmone, usava dati ottenuti da RANKED RESPONSE. Alla fine degli anni '60, Green e i suoi colleghi Vithala Rao, Frank Carmone e Arun Jain iniziarono numerosi esperimenti con il programma

MONANOVA: un documento di lavoro di Green e Rao apparso nel 1969 diventò il primo articolo di C.A..

A seguito di questo, Johnson [1974] e Westwood, Lunn e Beazley [1974] pubblicarono articoli circa “Johnson’s two attributes at a time trade off model”.

L’uso del metodo MONANOVA di Kruskal e di quello LINMAP PROGRAMS (nato come modello non metrico) di Shocker e Srinivasan [1977] portò ben presto alla consapevolezza che le variabili dipendenti del ratings based analizzate con dummy-variable regression techniques portavano ad una robusta alternativa alla “ordinally based data procedures” [Carmone, Green e Join, 1978; Cattin e Wittink, 1976].

‘Orthogonal main-effects plans’ si basa sul fractional factorial designs di Addelman [1962] che riduce in modo significativo l’onere conoscitivo dell’intervistato per rispondere alle domande sulle descrizioni dei prodotti nell’ambito del full profile.

Hence, nella metà degli anni ’70, capì che le condizioni erano mature per una veloce diffusione dei METODI METRICI DI C.A., con l’uso della “dummy variable regression”.

Nella metà degli anni ’80, Jonmson [1987] introdusse il suo “adaptive C.A. program” che usava il “Graded paired comparisons” quale set di input per il modello.

Circa negli stessi anni, Herman [1988] risponde con la presentazione di “PC BASED PACKAGES” che usano gli stimoli del full profile basati su disegni ortogonali. Entrambi i pacchetti contenevano simulatori congiunti, erano facili nell’utilizzo e moderatamente costosi.

Inizialmente vennero impiegati prettamente nelle aziende di ricerca commerciale, ma successivamente aprirono un largo e desideroso mercato di applicazioni nei campi più diversi.

Dalla fine degli anni ’70, la C.A., è diventata veramente all’avanguardia. Wittink e i suoi coautori Wittink e Cattin, 1989; Wittink, Vriens e Burhenne ,1994, svilupparono costose indagini che comprovavano la rapidità con cui questa tecnica era stata usata dai ricercatori e i suoi sviluppi teorici ed applicativi.

L’impulso che sta dietro alla diffusione della C.A. riflette: l’influenza e il contributo esercitati da studiosi e praticanti della metodologia, la disponibilità di software facili da usare e la credibilità evidenziata dalle applicazioni di business.

La C.A. in modo diverso dall'MDS e dal CLUSTERING, elargisce con i problemi centrali, misurando "buyer trade offs", al fine di sviluppare nuovi prodotti, riformulare prodotti vecchi e stagnanti e stimare le funzioni di prezzo-domanda.

MDS e CLUSTERING ANALYSIS sono spesso usate come tecniche ausiliarie per l'analisi dei dati in presentazione, in ogni caso, hanno avuto un remunerativo impiego negli studi di C.A. che coinvolgevano la "Buyer segmentation" [Green e Krieger, 1991] e la rappresentazione percentuale in maps dei risultati di C.A..

## **1.2 Sviluppo della Conjoint Analysis negli anni '80**

Dopo la seconda decade, gli sviluppi tecnologici della C.A. si sono rapidamente susseguiti, accompagnati e alimentati da migliaia di applicazioni (tabella 1). Durante gli anni '80, due sviluppi risultarono particolarmente interessanti, e riguardavano esplicitamente il modello di sviluppo e di applicazione:

- CHOICE BASED CONJOINT MODELS
- HYBRID CONJOINT MODELS, che includeva il Johnson's adaptive conjoint analysis model

In TRADITIONAL C.A., i rispondenti tipicamente valutano vari prodotti o profili di offerta, presentati uno alla volta, su una scala di probabilità d'acquisto.

In CHOICE BASED C.A., i rispondenti confrontano i profili di due o più espliciti competitori, che variano su uno o più livelli di attributi.

In questo caso, lo scopo è: scegliere il profilo preferito dal set di profili o in alternativa dare 100 punti tra tutti i profili del set, costruendo un'effettiva scala di preferenze relative.

Nella CHOICE BASED C.A., gli analisti impiegano usualmente un modello logico multinominale: "multinomial logit models, (MNL)" oppure occasionalmente il "probit-based models".

La TRADITIONAL C.A. è spesso utilizzata quando un prodotto nuovo entra in un nuovo e stabile mercato, in cui i competitori sono inesistenti o passivi in termini di risposta competitiva alle “new entry”.

Nella CHOICE BASED C.A., invece, gli analisti assumono i competitori attivi, cioè che possono modificare gli attributi dei loro prodotti, tra cui si possono avere modifiche di prezzo. I precursori della CHOICE BASED C.A. (che usavano l’MNL) erano studiosi di “econometrician”, e tra questi ricordiamo Mc Fadden [1974].

Il lavoro di quest’ultimo fu presto riconosciuto e usato da numerosi ricercatori di marketing inclusi Punj e Staelin [1978].

Gensch e Recker [1979] usarono questo modello per sviluppare un’alternativa alla regressione per analizzare il “CROSS SECTIONAL CHOICE DATA”.

Batsell e Lodish [1981] illustrarono il MULTINOMIAL LOGIT’S USE nel modellamento della scelta individuale sulla replicazione dei choice sets.

Il loro modello rese una quota di scelta prevedibile per le alternative in competizione nel “choice sets”.

Con l’aumento delle applicazioni della C.A. si sviluppò il desiderio di espandere il numero di attributi e di livelli che potevano essere composti. I modelli “HYBRID” impiegarono il “SELF EXPLICATED DATA COLLECTED” sia sulla desiderabilità che sull’importanza di attributi e livelli. I rispondenti, quindi, considerano un set più piccolo di profili da valutare, rispetto al Full Profile.

I primi modelli “HYBRID” usavano la “stagewise regression” per adattare e semplificare i modelli per l’applicazione del “SELF-EXPLICATED DATA”. Il valore di questi modelli ingannava nella grande accuratezza che raggiungeva, se confrontato con i “NON HYBRID MODELS”, per quanto riguardava la valutazione degli attributi.

Le risposte nel Full Profile essenzialmente servivano per affinare l’importanza degli attributi nel “SELF EXPLICATED”.

Più recentemente, Green e Krieger [1986] estesero i modelli “HYBRID” per permettere la stima dei parametri e dei livelli individuali. Ne risultarono quattro separati modelli di elevata generalità. Successivamente, l’obiettivo diventa quello di usare il “SELF-EXPLICATED DATA” in origine per stimare le “part worth” riguardanti i vari attributi, mentre l’uso del Full Profile Analysis serviva per produrre stime migliori di importanza dei vari attributi.

L'uso commerciale più importante dei “conjoint models” viene presentato da “Adaptive Conjoint Analysis Program (ACA)” di Johnson.

ACA è un modello “HYBRID” che contiene:

1. “SELF- EXPLICATED DATA” riguardante la desiderabilità e l'importanza di attributi e livelli.
2. “PAIRED COMPARISON”, cioè la presentazione di coppie di profili parziali (tipicamente consistono di attributi su due o forse tre livelli), presi dal Full Set di attributi. Al rispondente viene chiesto di scegliere tra gli elementi di ogni coppia e definire la percentuale di preferenza.

Gli sviluppi di ACA [SAWTOOTH SOFTWARE 1999] hanno continuato ad introdurre utili raffinamenti alla versione originale.

## GREEN, KRIEGER, WIND

---

Choice-based conjoint [McFadden 1974; Gensch and Recker 1979; Batsell and Lodish 1981; Mahajan, Green, and Goldberg 1982; Louviere and Woodworth 1983]  
Three-way multivariate conjoint analysis [DeSarbo et al. 1982]  
Number of attribute levels effects on derived conjoint importance [Wittink, Krishnamurthi, and Nutter 1982]  
Constrained parameter estimation in conjoint analysis [Srinivasan, Jain, and Malhotra 1983]  
Hybrid models for conjoint analysis [Green, Goldberg, and Montemayor 1981; Green 1984]  
Introduction in 1985 of Bretton-Clark's full-profile conjoint techniques [Herman 1988]  
Introduction in 1985 of Sawtooth Software's adaptive conjoint analysis [Johnson 1987]  
Factor analytic approaches to individualized conjoint analysis [Hagerty 1985]  
Conjoint analysis and MDS in tandem [Green, Krieger, and Carroll 1987]  
Reliability and validity testing [Bateson, Reibstein, and Boulding 1987]  
Simultaneous conjoint parameter estimation and segmentation [Kamakura 1988]  
Bretton-Clark's second generation, full-profile suite of programs [Herman 1988]  
Conjunctive-compensatory self-explicated models [Srinivasan 1988]  
Componential segmentation with optimization features [Green, Krieger, and Zelnio 1989]  
Compensatory model problems in negatively correlated environments [Johnson, Meyer, and Ghose 1989]  
New experimental designs for conjoint [Steckel, DeSarbo, and Mahajan 1991]  
A reservation price model for optimal pricing [Kohli and Mahajan 1991]  
A review of experimental choice analysis [Batsell and Louviere 1991]  
Latent class conjoint analysis [DeSarbo et al. 1992]  
Constrained part-worth estimation [van der Lans and Helsen 1992]  
Modeling hierarchical conjoint processes [Oppewal, Louviere, and Timmermans 1994]  
Concomitant variable latent class modeling [Kamakura, Wedel, and Agrawal 1994]  
Hierarchical Bayes models for conjoint analysis [Allenby, Arora, and Ginter 1995; Allenby and Ginter 1995; Lenk et al. 1996]  
A comparison of metric conjoint models [Vriens, Wedel, and Wilms 1996]  
Utility balanced experimental designs [Huber and Zwerina 1996]  
Competitive interaction simulators [Choi, DeSarbo, and Harker 1990; Green and Krieger 1997]  
Mixture models for segmentation [Wedel and Kamakura 1998]  
Commercial windows-based, choice-based conjoint [Sawtooth Software 1999]  
Krieger and Green's hybrid choice-based conjoint model [Vavra, Green, and Krieger 1999]  
Latent class conjoint analysis [Ramaswamy and Cohen 2000]  
Hierarchical Bayes applied to Internet recommendation systems [Ansari, Essegaiier, and Kohli 2000]  
Response latencies and conjoint analysis [Haaijer, Kamakura, and Wedel 2000]

---

**Tabella 1: questa è una lista parziale di contributi alla conjoint analysis nel periodo 1974-2000.**

### **1.3 Sviluppo della Conjoint Analysis negli anni '90**

Gli sviluppi di larga scala, nel '90, si basarono sui "HIERARCHICAL BAYESIAN MODELING" per la stima delle differenze individuali, rilevate dall'implementazione dei "CHOICE BASED MODELS".

Prima, le preferenze nei "CHOICE BASED MODELS" erano stimati dall'unione dei dati attraverso tutti gli individui o attraverso i "latent class methods", cioè una disgregazione parziale, applicata da De Sarbo e altri [1992] e Ramaswamy e Cohen [2000]. Il lavoro di Allenby, Arora e Ginger [1995], Allenby e Ginger [1995] e Lenk, Desarbo, Green e Young [1996], ha reso capaci gli utilizzatori della "CHOICE BASED C.A.", di ottenere i livelli individuali di stima delle "Part Worth" basata su "HIERARCHICAL BAYESIAN METHODS". Sawtooth software ha recentemente aggiunto questo tipo di modulo al suo "CHOICE BASED CONJOINT SOFTWARE".

I parametri individuali sono sia "SELF-CONSISTENT", che "COERENTI", ma diversi dall'"AGGREGATED DATA", in questo modo i dati individuali riceveranno più peso nella stima delle "part worths".

Recentemente, Vavra, Green e Krieger [1999] hanno proposto un altro approccio per la "CHOICE BASED C.A.". Hanno sviluppato un "HYBRID CHOICE BASED MODEL" che combina "SELF-EXPLICATED DATA" con le risposte nel FULL PROFILE.

Questo modello non richiede una media pesata dei dati dell'individuo riferita all'intero gruppo.

### **1.4 Applicazioni**

Dall'estesa lista (tabella 2) di applicazioni di C.A., due sono particolarmente interessanti: il cliente guida il progetto degli HOTELS DI MARRIOTT'S COURTYARD [Wind 1989], e il progetto e valutazione del New Jersey e New York EZ-PASS ELETTRONIC TOLL COLLECTION SYSTEM [Vavra, Green e Krieger, 1999].

1. HOTELS DI MARRIOTT'S COURTYARD: Agli inizi degli anni '80, MARRIOTT MANAGEMENT voleva progettare un'ottima catena di hotel di prima categoria per "BUSINESS TRAVELERS" i quali non avevano bisogno di molti servizi speciali, come scale mobili e 'marriott ehyatt', che furono omessi dallo studio, che invece contemplava il prezzo ed altra sette sfaccettature dell'Hotel:

- DECORI ESTERNI
- DECORI NELLE CAMERE
- SERVIZI RISTORANTE
- RIPOSO FACILITATO
- SERVIZI GENERALI
- ATTIVITA DI TEMPO LIBERO (fitness club)
- SERVIZI SPECIALI DI SICUREZZA

Gli analisti hanno sviluppato circa 50 attributi con un totale di 160 livelli. I modelli includevano "HYBRID CONJOINT C.A." e un primitivo tipo di "CHOICE BASED C.A."

Gli analisti usarono "computers simulations" per far vivere il servizio al cliente. Lo studio fa esteso uso di "VISUAL PROPS", cioè materiale visivo (figure e modelli tridimensionali), come stanze sperimentali in cui l'arredo e i decori erano sistematicamente variati secondo il progetto sperimentale.

Da tutti i conteggi svolti, il "COURTYARD STUDY" era un successo. Marriott implementò pressoché tutte le proposte del progetto e dopo estese l'approccio ad altri nuovi prodotti (per esempio Marriott Suites). I risultati furono inoltre usati nei progetti pubblicitari e di opuscoli del COURTYARD dal 1990. Nel 1990, Marriott aveva più di 300 Courtyards con impiegati più di 15000 persone, oggi, ci sono più di 450 Courtyards nel mondo intero con fatturati annui di miliardi di dollari.

Da allora Marriott ha applicato la C.A. in ogni progetto sperimentale "TIME SHARE" inerente alla partizione del tempo vacanziero, delle stanze e dei prezzi attrattivi.

2. Nell' "EZ-PASS TOOL COLLECTION PROJECT", nel New Jersey e New York venne realmente sviluppato un nuovo prodotto con lo scopo di rendere più rapido e semplice il passaggio dei veicoli sui ponti, nei tunnel e sulle strade pubbliche.

I pendolari usavano un cartellino elettronico (transponder) incollato all'interno del vetro del veicolo e quando si avvicinavano alla corsia di pedaggio, un'antenna leggeva i dati del veicolo e del cliente, memorizzando tutte le informazioni che poi venivano registrate nel cartellino. L'informazione era elettronicamente spedita on-line ad un computer che, valutando le informazioni, deduceva il pedaggio da addebitare al cliente.

Il progetto iniziò nel 1992, con due principali domande:

- COME SAREBBE STATO CONFIGURATO EZZ-PASS?
- CHE LIVELLO DI RISORSE SAREBBE STATO INVESTITO PER LA SUA IMPLEMENTAZIONE?

I due Stati condussero un grande studio di C.A. (con più di 3000 rispondenti), attraverso una procedura chiamata TELEPHONE-MAIL-TELEPHONE, nella quale i rispondenti qualificati ricevevano un pacchetto di materiale per rispondere e completare il questionario, e un video dimostrativo. Quest'ultimo conteneva un filmato di 11 minuti di informazioni che descrivevano il problema associato al sistema del pedaggio stradale e i benefici che l'EZZ PASS avrebbe introdotto. Gli analisti usarono sette attributi congiunti in relazione con tale problema:

- NUMERO DI CORSIE DISPONIBILI
- ITER BUROCRATICO PER L'ACQUISIZIONE DEL CARTELLINO
- COSTI PER L'ACQUISIZIONE
- PREZZO DI PEDAGGIO
- FATTURAZIONE
- METODOLOGIA DI PEDAGGIO
- ALTRI USI DEL CARTELLINO

Il progetto di studio analizzava i dati individuali dei rispondenti e di tutto il campione, inerente all'età del guidatore, al livello del campione analizzato, alla zona ed ai mezzi. Complessivamente (all'equilibrio) le previsioni fatte nel 1992 dalla raccolta dei dati era un utilizzo pari al consueto 49%. Dopo sette anni la percentuale si era ridotta al 44%; nel futuro l'uso era previsto circa del 49%.

Entrambi i progetti "COURTYARD BY MARRIOTT" e "EZ-PASS" mostrano l'abilità della C.A. a guidare e perseguire scoperte che portano a ridisegnare le caratteristiche del CUSTOMER-DRIVEN, gli USI DEI CONSUMATORI e le PREVISIONI DI VENDITA.

---

AT&T's first cellular telephone—Chicago-based study of 1,000 drivers' reactions to cell phone features of the new "honey-comb" relay system

Ford Fairlane—conjoint analysis used in redesign of Ford to reflect general automotive downsizing objectives

IBM RISC 6000 workstation—conjoint used to measure potential buyer reactions to variations in performance and reliability features of a new workstation

Squibb's captopril antihypertensive—six-country study of physicians' evaluations of Capoten's efficacy and safety features

Tagamet (SKF) and Zantac (Glaxo) ulcer drugs—competitive pricing and analysis of demand elasticities

Health maintenance plans—study conducted by the American Association of Retired Persons; results submitted to Congress

US Navy benefit packages for reenlistment—conjoint used to develop menu of new reenlistment plans based on individual differences in types of duties, health needs, and sign-over bonuses

Antidumping litigation—AT&T vs. Pacific-rim manufacturers legal dispute regarding small business telephone equipment; case adjudicated in AT&T's favor

Continental vs. American Airlines litigation—conjoint study of travel agents' trade-offs among airline flight selections

FedEx new services study—trade-off study of customer reactions to new methods for tracking delayed and lost letters and packages

Intermittent windshield wipers litigation—study to determine consumer evaluations of the derived "willingness to pay" for the intermittent wiper feature

MasterCard and Diner's Club—new travel and entertainment features evaluated for each case

Monsanto's herbicide packaging study—consumer reactions to advanced packaging devices for liquids, solids, and aerosols

Polaroid's instant camera design—study of consumer reactions to new features and camera design aesthetics

Marriott time-share units—development of optimal interior and exterior décors, services, and price

Ritz Carlton—conjoint used to develop hotel décor and services

Japanese cable TV—conjoint survey of Japanese consumers' trade-offs among services and prices of satellite TV

Shell and Texaco's brand equity—conjoint study of price, brand, and alternative company positions for the Shell and Texaco Corporations

UPS services study—conjoint used to examine customers' evaluations of four major suppliers of overnight letter and package delivery

---

**Tabella 2: questa tabella illustra le applicazioni di conjoint analysis che sono state condotte dagli autori e dai loro collaboratori.**

## 1.5 Prospettive future

Dopo 30 anni di sviluppi e applicazioni, la C.A. sembra aver superato il test del tempo. Mentre nuovi passi avanti possono essere meno frequenti, il metodo è in continua crescita nella profondità e larghezza d'utilizzo.

Si aspettano di vedere successivi e ulteriori sviluppi, magari inerenti a:

1. Nuovi "Simulator-optimizers" che possono massimizzare ritorni finanziari o quota di mercato [Vavra, Green e Krieger, 1999].
2. Nuove classi di problemi, che includono selezioni da liste specifiche e collegamenti tra modelli inerenti alle telecomunicazioni e ai servizi bancari [Ben-Akiva e Gershensfeld, 1998].
3. Immagini più realistiche per descrivere livelli degli attributi, per esempio, usando "Virtual Reality Displays".
4. Espansione continua di applicazioni di C.A. in ogni campo come: turistico, spettacolo/intrattenimento, mantenimento della salute, gioco d'azzardo e dispute legali.
5. Nuove applicazioni, come la recente implementazione su internet, che include 'activebuyersguide.com', 'personalogie.com' e 'conjointonline.com', siti che tipicamente usano modelli ibridi di C.A. per ricavare le preferenze del compratore per la merce acquistabile via web [Ansari, Essegai e Kohl, 2000].
6. Studi aggiunti di conjoint di affidabilità e validità [Haaijer, Kamakura e Wedel, 2000, Vriens, Wedel e Wilms, 1996].
7. Nuovi dinamici "CONJOINT SIMULATORS" che considerano sequenze competitive di AZIONE/REAZIONE [Choi, Desarbo e Harker, 1990, Green e Krieger, 1997].
8. Prototipi Simulators (per esempio, test cards) che permettono agli analisti di misurare le preferenze dei rispondenti in un ambiente realistico.

In breve, malgrado la sua maturità, la C.A. è ancora distante dall'inattività, perché i metodi trattano di dilaganti problemi inerenti le preferenze e le scelte del compratore, una questione sempre attuale, quindi il futuro promette uno sviluppo ed applicazione

continua. In tutto e per tutto la C.A. e le tecniche precedenti di CLUSTER ANALYSIS E MDS devono il loro successo e la loro grande diffusione alla disponibilità di software poco costosi e facili all'uso.

All'inizio, BELL LABORATORIES e ISTITUTI DI MARKETING SCIENCE, giocano un ruolo importante nel mercato "MAINFRAME SOFTWARE" disponibile in diverse versioni per usi sia industriali che accademici.

Con la crescita dei personal computer, il SAWTOOTH SOFTWARE di Johnson e il BRETTON- CLARK COMPANIES di Herman crearono pc software per usi di business e versioni accademiche affinché attraverso la continua applicazione e ricerca si riuscisse ad acquisire maggior risolutezza e precisione. SAWTOOTH SOFTWARE ha mantenuto i contatti con business ed accademia attraverso nuove lettere, incontri annuali e continui sviluppi di software [SAWTOOTH SOFTWARE 1999], per implementare nuove metodologie di ricerca, come il CHOICE BASED MODELS.

Il forum annuale ADVANCED RESEARCH TECHNIQUES dell'AMERICAN MARKETING ASSOCIATION prevedeva ogni anno un incontro per un fruttuoso scambio di idee tra accademici e praticanti.

Dal suo inizio, la C.A. ha sviluppato metodologie sulla base di nozioni nate da matematici, psicologi, psicometrici, statistici, economisti e ricercatori.

Queste idee riguardavano la progettazione sperimentale, la stima parametrica, la descrizione dei modelli già costruiti, le normative degli stessi e la valutazione comparata dei vari modelli validi disponibili.

Le conseguenze delle applicazioni pratiche attestarono il valore e la stabilità della C.A. nel tempo. Tutto questo è dovuto all'azione reciproca dei teorici e dei praticanti che collaborarono insieme per migliorare questa importante e preziosa tecnica statistica [Gustaffsson, Hermann e Huber, 2000].

La CONJOINT METHODOLOGY continua a crescere, così come accademici e praticanti imparano utili cose gli uni dagli altri, in uno scambio reciproco di dati e valutazioni utili.

## 1.6 Riflessioni e prospettive

Thomas Saaty [1980] aveva introdotto un approccio per la misura dell'utilità in presenza di attributi multipli, chiamato Analytic Hierarchy Process [AHP]. Questo insieme ad altri metodi furono applicati da "operations researchers" e "managements scientists" (OR/MS) per risolvere importanti problemi di management e come soluzione a decisioni di mercato. OR/MS avevano meno familiarità rispetto ad un altro metodo, la Conjoint Analysis, che ormai viene applicato da più di trent'anni, principalmente da marketing and business researchers.

La Conjoint Analysis (COA) fa parte della famiglia dei "Modelli di preferenza Multiattributivi" che consentono di esprimere le preferenze di singoli consumatori per un determinato tipo di prodotto attraverso la specificazione di alternative multiattributive dello stesso, nel senso che ogni preferenza può essere intesa come funzione della posizione di un particolare stimolo su di un numero di scale relative agli attributi usati.

Tutto ciò può essere sintetizzato, per un generico individuo  $d$ -esimo, dal seguente modello:

$$y_{j(d)} = F_{j(d)}(f_{j1(d)}, f_{j2(d)}, \dots, f_{jk(d)})$$

con:

$y_{j(d)}$  : misura delle preferenze che lo stimolo  $j$  ( $j = \{1, 2, \dots, n\}$ ) assume per l'individuo  $d$  ( $d = \{1, 2, \dots, p\}$ );

$F_{j(d)}$  : funzione di preferenza dell'individuo  $d$ ;

$f_{ji(d)}$  : modalità dell'attributo  $i$  ( $i = \{1, 2, \dots, k\}$ ) percepita dall'individuo  $d$ ;

La Conjoint Analysis si è sviluppata dalla "seminal research" di Luce and Tukey [1964], e questo primo contributo teorico fu utilizzato da molti "psychometricians", tra cui Carroll [1969], Kruskal [1965] e Young [1969].

Queste ricerche svilupparono un tipo di modello non metrico per misurare le utilità parziali, "part-worths", delle preferenze dei rispondenti (valori dei livelli degli attributi), ordinate attraverso stimoli multiattributo come descrizione di prodotti o servizi.

La Conjoint Analysis è, di gran lunga, il metodo più usato di ricerca di mercato per analizzare i “consumer trade offs”, cioè i compromessi che il consumatore deve sostenere ogni giorno nelle sue scelte.

Un’indagine condotta da Wittink e Cattin [1989] e Wittink, Vriens e Burhenne [1994] ne prova la sua diffusione a livello mondiale.

Non è difficile comprendere perchè le ricerche e le applicazioni di questo metodo si siano sviluppate così rapidamente: la C.A. parte da una domanda centrale del management, **perché il consumatore sceglie di acquistare una certa marca di prodotto e non un'altra?** La C.A. cerca di risolvere situazioni di mercato in cui il consumatore, nelle sue scelte, deve considerare la variazione simultanea di due o più attributi del prodotto o del servizio.

Il problema delle decisioni di mercato è un compromesso sulla possibilità che: l’opzione X è migliore dell’opzione Y su un attributo A mentre Y sia migliore di X sull’attributo B e altre varie estensioni di questi conflitti. La C.A. rappresenta le decisioni quotidiane del consumatore: che marca di tostapane, di automobile o di macchina fotocopiatrice comprare o affittare?

La metodologia attraverso una ricerca di mercato è in grado di raccogliere informazioni sui trade-offs di centinaia e migliaia di rispondenti, misurarli, analizzare le preferenze dei rispondenti e le intenzioni d’acquisto: serve per simulare come i consumatori reagiscono ai cambiamenti nei prodotti o all’introduzione di nuovi prodotti all’interno di un segmento competitivo esistente.

I ricercatori usano la C.A. per prodotti industriali, di consumo, servizi e offerte no-profit.

Per capire i concetti base della C.A., si assume che un venditore di carte di credito voglia esaminare la possibilità di modificare la linea attuale dei servizi offerti. Uno dei primi step nel progetto di un “conjoint study” è lo sviluppo di un set di attributi, e dei corrispondenti livelli che caratterizzano la sfera competitiva. Focus groups, interviste approfondite ai consumatori e gruppi interni di esperti, sono alcune fonti usate dai ricercatori per creare il set di attributi e livelli che devono guidare il resto dello studio.

In uno studio recente di fornitori di carte di credito, i ricercatori hanno usato 12 attributi con un minimo di 2 a un massimo di 6 livelli ognuno, per un totale di 35 livelli (tabella 3); in ogni caso il totale delle possibili combinazioni o profili è 186,624.

Gli esperti di C.A. utilizzano quasi esclusivamente vettori ortogonali [Addelman, 1962] e altri tipi di piani fattoriali frazionati che permettono di ridurre il numero delle combinazioni descritte, per cui il rispondente vede solo una frazione del numero totale dei profili. In questo problema, un vettore di 64 profili (meno dello 0.04% del totale) è sufficiente per stimare tutti gli effetti principali dei livelli degli attributi su base non correlata.

Nel progetto di studio si è utilizzato un tipo di C.A. chiamato Hybrid Conjoint Design [Green e Krieger, 1996], e ogni rispondente ha risposto considerando solo otto profili tra i sessantaquattro iniziali.

## GREEN, KRIEGER, WIND

Annual price (\$)					
0	10	20	50	80	100
Cash rebate (end-of-year, on total purchases)					
None		$\frac{1}{2}\%$		1%	
800 number for message forwarding					
None	9-5 weekdays		24 hours per day		
Retail purchase insurance					
None		90 days' coverage			
Common carrier insurance (death, injury)					
None		\$50,000		\$200,000	
Rental car insurance (fire, theft, collision, vandalism)					
None		\$30,000			
Baggage insurance					
None		\$2,500 depreciated cost		\$2,500 replacement cost	
Airport club admission (based on small entrance fee)					
No admission		\$5 per visit		\$2 per visit	
Card acceptance					
Air, hotel, rental cars (AHC)					
AHC and most restaurants (AHCR)					
AHCR and most general retailers (AHCRG)					
AHCR and department stores only (AHCRD)					
24-hour medical/legal referral network					
No		Yes			
Airport limousine to city destination					
Not offered		Available at 20% discount			
800 number for emergency car service					
Not offered		Available at 20% discount			

**Tabella 3:** questi attributi (e livelli di attributi) descrivono il set di servizi potenziali che si vorrebbero offrire a chi sottoscrive una carta di credito.

Per ogni studio, i ricercatori preparano delle “prop cards” o cartellini (tabella 3); i rispondenti segnano la carta in termini di preferenza e successivamente ogni carta è ordinata su una scala da zero a cento di probabilità di acquisto.

Nel caso in cui l'applicazione di C.A. riguardi pochi attributi, per esempio sei/sette con ognuno due/tre livelli, i rispondenti ricevono tutti i profili: questa metodologia prende il nome di Full Profile, e i cartellini somministrati variano da 16 a 32 considerando l'esempio. In questo caso i rispondenti separano le "prop cards" in quattro/otto categorie prima di ordinarle per probabilità di acquisizione riguardante ogni profilo all'interno di ogni categoria.

I quattro tipi più importanti di procedure per la raccolta di dati, usati quotidianamente per la Conjoint Analysis sono:

1. FULL PROFILE TECHNIQUES: ogni rispondente riceve un set completo di profili "prop cards". Dopo aver suddiviso le cards in categorie ordinate, il rispondente stima ogni carta su una scala di probabilità d'acquisto da 0 a 100.
2. IN COMPOSITIONAL TECHNIQUES: come il "Casemap procedure" [Srinivasan, 1988], dove ogni rispondente stima la desiderabilità di ogni set di livelli di attributi su una scala da 0 a 100 e ordina gli attributi su una scala di importanza. Questo approccio è tipicamente chiamato SELF-EXPLICATED PREFERENCE DATA COLLECTION.
3. HYBRID TECHNIQUES: ogni rispondente compie un SELF-EXPLICATED EVALUATION TASK e valuta un subset di cards del full profile [Green, Goldberg, e Montemayar]. La risultante funzione di utilità è una composizione di dati ottenuta da entrambi i compiti.
4. ADAPTIVE CONJOINT ANALYSIS: è un HIBRID TECHNIQUE sviluppata dalla Sawtooth Software [Johnson 1987], dove ogni rispondente prima svolge una "SELF EXPLICATED TASK" e poi valuta un set parziale di profili proposti due alla volta. I profili parziali consistono usualmente in due/tre attributi per stimolo (cards). I ricercatori variano i profili parziali, in quanto dipendono dalle risposte precedenti e dal confronto precedente (PAIRED COMPARISONS). Il rispondente valuta i profili parziali prendendoli due alla volta su una graduatoria o scala di confronto. Sia il SELF EXPLICATED

TASK, che il PAIRED COMPARISONS sono amministrati dal computer [Johnson, 1987].

1	2
Annual price \$20	Annual price \$50
Cash rebate None	Cash rebate ½ %
800 number for message forwarding None	800 number for message forwarding 24 hours per day
Retail purchase insurance None	Retail purchase insurance None
Common carrier insurance \$50,000	Common carrier insurance None
Rental car insurance \$30,000	Rental car insurance \$30,000
Baggage insurance \$2,500 depreciated cost	Baggage insurance None
Airport club admission \$2 per visit	Airport club admission \$5 per visit
Card acceptance Air, hotel, rental cars	Card acceptance Air, hotel, rental cars and most restaurants
Medical-Legal No	Medical-Legal Yes
Airport limousine Not offered	Airport limousine Available at 20% discount
800 number for emergency car Available at 20% discount	800 number for emergency car Not offered

Figura 1: Queste prop cards illustrano tutti i possibili servizi specifici che una carta di credito può offrire. Per ogni card, il rispondente, indica con quale preferenza sottoscriverebbe la carta di credito su una scala da 0 a 100.

## 1.7 I modelli della Conjoint Analysis

Gli analisti stabilirono che quello che era conosciuto come “**part worth model**” rappresentava il giudizio valutativo dei rispondenti, e si poteva ottenere usando i full profile, il self-explicated or hybrid approches.

I modelli principali suggeriti da Green e Srinivasan sono: a) modello *vettore*, lineare; b) modello *punto-ideale*, lineare-quadratico; c) modello *parth worth*, discontinuo.

Il **modello vettore** esprime la preferenza  $y_j$  (utilità globale) per il prodotto  $j$  nel seguente modo:

$$y_j = \sum_{i=1, \dots, k} w_i f_{ij} .$$

dove  $w_i$  rappresenta il peso dei  $k$  attributi e  $f_{ij}$  il livello dell' $i$ -esimo attributo relativo al  $j$ -esimo stimolo. La relazione che lega variabile dipendente ed indipendenti è di tipo lineare: la preferenza complessiva di ciascun profilo è legata linearmente ai valori degli attributi per meno dei coefficienti di ponderazione  $w_i$ . Data la sua forma funzionale, il modello vettoriale risulta adeguato per una conjoint analysis in cui il livello di gradimento di un bene o servizio cresce al crescere dell'attributo ( $w_i > 0$ ) o diminuisce al diminuire dello stesso ( $w_i < 0$ ): in presenza, quindi, di preferenza monotona e lineare. Il numero di parametri da stimare è il minore tra i tre modelli qui sopra riportati: corrisponde alla numerosità  $n$  degli attributi considerati nell'analisi. Il ridotto numero di coefficienti può accrescere il potere predittivo della C.A..

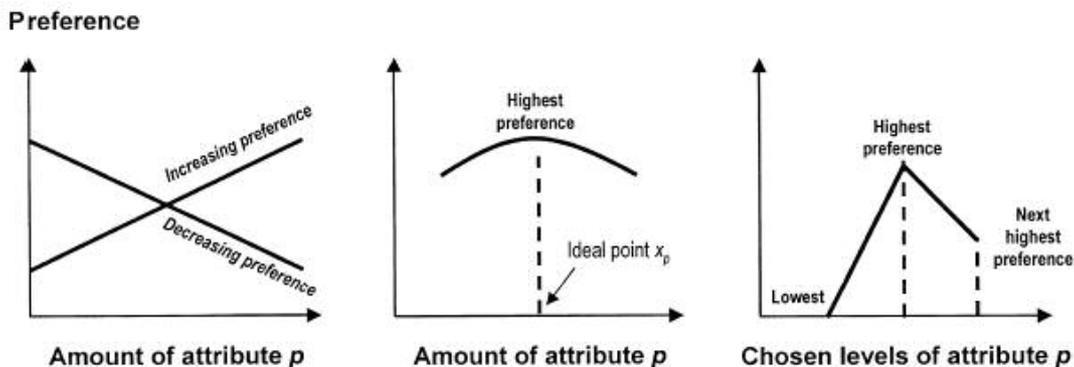


Figura 2: questi diagrammi illustrano il significato di preferenze lineari, preferenze riferite al punto ideale, e discrete (part-worth) preferences. Il terzo grafico mostra tre part-worths. Fonte: Green e Srinivasan [1978].

Il modello **punto-ideale** prevede che l'utilità complessiva di ciascun profilo sia correlata negativamente con la seguente misura (distanza del profilo  $j$ -esimo):

$$d_j^2 = \sum_{i=1, \dots, k} w_i (f_{ij} - o_i)^2 .$$

Per ogni attributo  $i$  esiste un livello ideale che rappresenta la scelta ottimale per l'individuo: supponendo che tutti i livelli selezionati per l'analisi siano egualmente possibili, il modello punto ideale afferma che, con certezza, l'individuo sceglie sempre ed unicamente il livello "ideale". Il modello ipotizza che non esistano differenze legate alla direzione di spostamento dal punto ideale dato che viene considerato il valore al quadrato della distanza: in presenza di un attributo con tre livelli, graficamente, la relazione appare come un semicerchio il cui punto più alto rappresenta il valore ideale. Questo tipo di modello trova largo uso quando si hanno attributi qualitativi associati a fattori sensoriali quali il gusto o l'odore: un sapore troppo dolce o troppo amaro viene scartato, mentre sarà preferito un giusto livello di dolcezza (livello ideale). Infine il modello **part-worth** ammette la forma funzionale più generale, la dipendenza tra la variabile dipendente e le indipendenti non ha limitazioni:

$$y_j = \sum_{i=1, \dots, k} f_i(x_{ij}).$$

dove  $f_i$  è la funzione rappresentante l'utilità del livello assunto dall' $i$ -esimo attributo. Il modello richiede che i livelli di ciascun attributo siano definiti design matrix da una distinta colonna di variabili dummy: data la presenza dell'intercetta, per stimare  $r$  (numero di livelli per ciascun attributo), si richiede l'uso di  $(r-1)$  dummies. Complessivamente i parametri da stimare saranno quindi  $n \cdot (r-1)$ . Questo modello è più flessibile e di carattere più generale degli altri due.

In figura 3, mostriamo le "part worth" per i livelli degli attributi descritti in tabella 3.

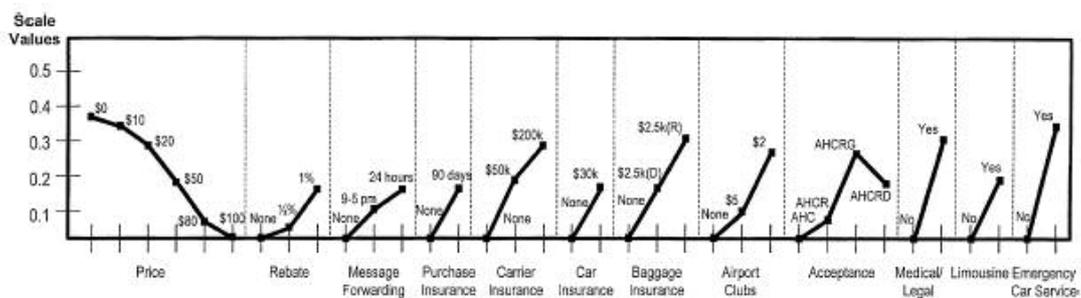


Figura 3: questo grafico illustra i valori delle part-worth valutati sulla scala presentata che dipendono dal prezzo, messaggi forwarding,.....servizio di emergenza per la macchina. Per esempio, il prezzo preferito è \$ 0.

Le “parth worths” sono spesso rappresentate in scala così che il livello più basso, senza ogni attributo, è zero.

Parlando in modo rigoroso, gli analisti valutano le “parth worth functions” i livelli discreti per ogni attributo; in ogni modo in molte applicazioni, gli analisti vanno ad interpolare tra i livelli di attributi continui, come il prezzo, quando si è all’interno delle “parth worth” nelle simulazioni di scelta del consumatore.

La rappresentazione in scala (assi verticali) è comune a tutti gli attributi, questo permette all’analista di sommare le “parth worth” di ogni livello dell’attributo per ottenere l’utilità totale del prodotto o servizio, cioè l’utilità di ogni profilo ottenibile dalla combinazione di livelli degli attributi.

## **1.8 Presentazione degli stimoli**

Nella raccolta dei dati, gli analisti enfatizzano spesso il FULL PROFILE e le PROCEDURE IBRIDE, incluso il Sawtooth’s Adaptive Conjoint Analysis.

Oltre all’impiego di paragrafi descrittivi di livelli di attributi, in alcuni studi industriali, spesso usano “profile cards” con una descrizione molto concisa dei livelli degli attributi (figura 1). Gli analisti aumentano l’uso di materiale figurativo, e grazie a questo supporto il compito del rispondente diventa più interessante e permette il trasferimento di informazioni con bassa ambiguità. Ultimamente i metodi “Conjoint” hanno avuto un incremento nelle applicazioni per la progettazione di prodotti fisici (per esempio, cibi, bevande, fragranze prodotti per la pulizia personale, personal care products), dove i ricercatori usano prototipi sperimentali virtuali.



## Capitolo 2

### ***CONJOINT ANALYSIS: FULL PROFILE***

#### ***2.1 Introduzione***

L'impiego di metodi statistici per la valutazione e il miglioramento della qualità dei sistemi di produzione di beni e servizi ha visto negli ultimi anni un consolidamento crescente come parte integrante dei sistemi di gestione aziendali. In un mercato sempre più complesso caratterizzato da elevata competitività, alto sviluppo tecnologico, volumi di produzione più elevati, maggiori e più differenziate esigenze individuali dei consumatori, l'orientamento aziendale alla qualità, diventa sempre più una strategia importante per aumentare la produttività e la penetrazione nel mercato, per raggiungere un forte vantaggio competitivo.

In tale contesto, il concetto di qualità, da un significato iniziale di conformità del prodotto a specifiche prettamente di tipo tecnico imposte internamente dall'azienda, ha gradatamente assunto un significato più ampio investendo tutti gli aspetti dell'attività aziendale (formazione del personale, progettazione, scelte economiche), fino a coinvolgere aspetti esterni all'azienda quali le esigenze e i bisogni dei clienti, assumendo quindi anche un significato di capacità di soddisfare le aspettative dei consumatori.

L'evoluzione del concetto di qualità di un bene o servizio ha posto quindi l'esigenza di affiancare ai metodi tradizionali del controllo statistico della qualità, altre metodologie statistiche tra cui i metodi per la valutazione e l'interpretazione di come i consumatori

percepiscano e valutino la qualità di prodotti o servizi indicata con il termine di “customer satisfaction”. Tra i contributi metodologici più innovativi in questo ambito, si situa la Conjoint Analysis (C.A.) finalizzata allo studio dei modelli di scelta dei consumatori a partire dai giudizi di preferenza espressi da questi ultimi relativamente a diverse possibili configurazioni dei prodotti o servizi. Le tradizionali procedure di analisi della “customer satisfaction, generalmente, prevedono un’indagine campionaria basata su un questionario in cui ciascuna domanda mira a valutare la soddisfazione del consumatore rispetto ad una specifica caratteristica del prodotto/servizio. A partire da queste valutazioni parziali si procede successivamente in fase di analisi dei dati raccolti alla stima della valutazione globale sull’insieme delle caratteristiche del prodotto/servizio. Al contrario la Conjoint Analysis prevede un’indagine campionaria in cui si sottopongono agli intervistati diversi profili del prodotto/servizio, ovvero diverse versioni di alternative di prodotto, definite in base a diverse combinazioni delle modalità di una serie di caratteristiche rilevanti del prodotto. Compito dei rispondenti è quello di assegnare a ciascun profilo il punteggio “globale” di gradimento o di redigere una graduatoria di preferenza degli stessi.

Sulla base del punteggio globale assegnato da ciascun intervistato ai vari profili di prodotto, in fase di analisi dei dati, sono stimate le preferenze “parziali” associate a ciascuna modalità di ogni caratteristica. La C.A. consente, quindi, di affrontare lo studio della soddisfazione e del comportamento di scelta del consumatore con un’ottica in cui si ritiene che il processo di formazione individuale delle preferenze sia di tipo “multiattributo”, ossia il consumatore crea, e quindi esprime, una preferenza sul prodotto considerandolo nella sua interezza cioè dal punto di vista di tutte le sue caratteristiche (attributi) “congiuntamente” considerati.

I risultati di un’analisi di C.A. possono essere utilizzati come supporto per lo sviluppo di nuovi prodotti/servizi, o per simulare gli effetti sulle preferenze dei consumatori di cambiamenti di alcune caratteristiche di prodotti già esistenti.

L’applicazione della C.A. può inoltre costituire un’integrazione alle tecniche di segmentazione del mercato, consentendo di individuare ad esempio profili dei consumatori che risultano maggiormente sensibili a particolari tipi di offerte di prodotto. In particolare in questo capitolo la procedura che verrà trattata per la raccolta di dati sarà la FULL PROFILE.

## 2.2 I concetti fondamentali e le fasi di un esperimento di Conjoint Analysis

Il disegno e l'esecuzione di un esperimento di C.A. si articolano secondo una serie di fasi che partendo dal problema di ricerca possono essere sinteticamente schematizzate nel modo seguente:

1. **Definizione dei fattori sperimentali e dei livelli**: in questa fase vengono individuate le caratteristiche del prodotto/servizio dette "attributi" o "fattori sperimentali" e per ciascun fattore vengono specificate le modalità o "livelli" di interesse. Le possibili combinazioni di livelli dei fattori definiscono i "profili" del prodotto/servizio che saranno sottoposti agli intervistati.
2. **Definizione del piano fattoriale e dei profili di prodotto**: sulla base dei fattori e dei livelli individuati al punto precedente viene definito il piano fattoriale che potrà includere solo gli effetti principali dei fattori selezionati (modello additivo) o anche le interazione tra i fattori (modello con interazione). La procedura di C.A. prevede usualmente la realizzazione di uno studio pilota con tutti i fattori su due livelli considerando un piano fattoriale completo che include le interazioni di ogni ordine. Successivamente l'analisi può essere ulteriormente riproposta considerando gli effetti principali veramente influenti sulla risposta sperimentale e le eventuali interazioni significative. Si consideri l'esempio di un'azienda interessata ad effettuare un esperimento di C.A. su un personal computer portatile. Vengono individuati tre attributi o fattori, ciascuno su due livelli.

FATTORE A	LIVELLO 1: 1,5 Kg LIVELLO 2: 2 Kg
FATTORE B	LIVELLO 1: 2 ore LIVELLO 2: 4 ore

FATTORE C	LIVELLO 1: 1000 EURO
	LIVELLO 2: 2000 EURO

Indicando i livelli bassi e i livelli alti dei tre fattori rispettivamente con “-“ e “+”, si può illustrare graficamente il piano fattoriale che è del tipo  $2^3$  ipotizzando di essere interessati ai soli effetti principali dei tre fattori:

PROFILI	FATTORI		
	A	B	C
1) a b c	+	+	+
2) a b	+	+	-
3) a c	+	-	+
4) a	+	-	-
5) b c	-	+	+
6) b	-	+	-
7) c	-	-	+
8) (1)	-	-	-

Ogni riga della matrice del piano definisce un “profilo”, ossia una combinazione dei livelli dei fattori selezionati (chiamato anche cartellino). In corrispondenza del piano  $2^3$ , risultano definiti otto profili, dove ciascun profilo è stato indicato nella figura con lettere minuscole in modo tale che se una lettera compare, allora il corrispondente valore è presente in quel profilo al valore alto, se è assente il fattore è presente al livello basso.

Il profilo in cui tutti i fattori sono presenti al livello basso è indicato con (1). Ciascun profilo corrisponde ad una diversa configurazione del prodotto. Ad esempio il profilo “3) a c” corrisponde ad un personal computer portatile con peso pari a 2 Kg, con durata di batteria pari a 2 ore e di costo pari a 2000 EURO. In generale dati  $k$  fattori, ciascuno caratterizzato da  $p_1, p_2, \dots, p_k$  livelli, il numero totale di profili del prodotto sarà pari a  $M = \prod p_k$ .

3. **Definizione della modalità di presentazione dei profili agli intervistati:** i profili, definiti al punto 2), sono sottoposti al giudizio di un campione di

rispondenti detti anche “valutatori”. La somministrazione dei profili ai rispondenti può essere effettuata in diverso modo. In particolare si parla di TRADITIONAL CONJOINT ANALYSIS o FULL PROFILE CONJOINT ANALYSIS quando ai valutatori sono sottoposti tutti i profili completi delle varie configurazioni del prodotto sotto forma generalmente di cartellini. Riprendendo l’esempio precedente, i cartellini possono essere rappresentati così:

CARTELLINO 1
PESO: 2 Kg DURATA: 4 ore PREZZO: 2000 EURO

CARTELLINO 8
PESO: 1 Kg DURATA: 2 ore PREZZO: 1000 EURO

I rispondenti daranno un giudizio in forma di punteggio o graduatoria relativamente all’insieme completo dei profili. Si parla invece di ADAPTIVE CONJOINT ANALYSIS quando, a differenza dell’approccio tradizionale, nella presentazione a coppie di profili successivi ai rispondenti, si tiene conto delle preferenze precedentemente ottenute, “adattando” di conseguenza i confronti successivi.

Nel seguito della trattazione si farà riferimento alla tipologia FULL PROFILE della CONJOINT ANALYSIS.

4. **Definizione della risposta sperimentale**: la preferenza dei rispondenti sui vari profili può essere espressa mediante punteggio (rating) su una scala metrica, ad esempio da 1 a 10, oppure mediante graduatoria di preferenza dei profili (ranking). Nel primo caso la variabile risposta è considerata continua in quanto

rappresenta il punteggio espresso su scala metrica e si parla quindi di Conjoint Analysis "Metrica"; nel secondo caso la variabile risposta è di tipo ordinale in quanto rappresenta la posizione di graduatoria e si parla di Conjoint Analysis "Non Metrica". Nel seguito si farà riferimento alla C.A. Metrica.

5. **Stima delle preferenze parziali dei livelli dei fattori e dell'importanza relativa di ciascun fattore:** dopo aver somministrato i profili di prodotto alle unità sperimentali (soggetti rispondenti), vengono stimate le preferenze parziali di ciascun livello dei fattori del prodotto ed una misura dell'importanza di ciascun fattore per ogni rispondente. Nella C.A. Metrica la stima può avvenire ipotizzando per la variabile risposta un modello di regressione lineare multipla, e stimando i parametri di tale modello con il metodo dei minimi quadrati.
6. **Verifica di ipotesi sugli effetti dei fattori del prodotto/servizio:** l'ultima fase della C.A. è relativa ai metodi di determinazione degli effetti maggiormente influenti sulla risposta sperimentale. Le procedure utilizzate in questa fase possono essere classificate principalmente secondo due diverse tipologie:
  - Nel caso in cui il piano sperimentale che ha prodotto l'indagine di C.A. sia di tipo non replicato (ovvero nel caso di un singolo rispondente o nel caso di aggregazione di giudizi di preferenza tramite la preferenza media) non essendovi gradi di libertà nella stima della varianza dell'errore sperimentale non è possibile condurre test di tipo parametrico. Quindi un metodo proposto in letteratura è quello introdotto da Daniel nel 1959 denominato NORMAL PROBABILITY PLOT che è un metodo grafico per la determinazione degli effetti che si discostano in modo sufficientemente "evidente" dal grafico di una distribuzione normale di media nulla. Tale approccio verrà delineato successivamente.
  - Nel caso in cui il piano sperimentale sia di tipo replicato, ovvero i rispondenti possono essere ragionevolmente identificati come appartenenti ad un gruppo omogeneo dal punto di vista delle variabili confondenti che potrebbero influire sul giudizio di preferenza, si può adattare la usuale procedura di ANALISI DELLA VARIANZA per piani di tipo multifattoriale.

### 2.3 Specificazione e stima del modello di utilità individuale

La C.A. pone in corrispondenza biunivoca il concetto di “preferenza” con quello di “utilità”, nel senso che il profilo di prodotto preferito è anche quello dal quale il rispondente ricava maggiore utilità. La “preferenza utilità” di un acquirente è considerata funzione delle modalità (livelli) degli attributi (fattori) rilevanti del prodotto/servizio in studio. Tale funzione viene detta “funzione di utilità individuale”. Tra i vari modelli di utilità individuale proposti in letteratura vi è il modello “additivo a coefficienti separati” secondo il quale l’utilità “globale” di un generico profilo di prodotto è data dalla somma delle utilità “parziali” di ogni livello dei diversi attributi che definiscono il prodotto.

Dati  $k$  fattori, ciascuno caratterizzato da  $p_1, p_2, \dots, p_k$  livelli, sia  $M = \prod_{k=1 \dots k} p_k$  il numero dei profili del prodotto. Per ciascun profilo costruiamo una variabile binaria come segue:

$d_{mkp}$	1 se il profilo $m$ presenta l’attributo $k$ con livello $p$
	0 altrimenti

dove  $m$  indica il generico profilo, con  $m = 1, \dots, M$ ;  $k$  indica il generico fattore, con  $k = 1, \dots, K$ ;  $p$  indica il generico livello del fattore  $k$ , con  $p = 1, \dots, P_k$ .

La funzione di utilità globale per il profilo  $m$ , con riferimento ad un generico rispondente, si può esprimere secondo il seguente modello:

$$(1) \quad Y_m = \sum_{k=1} \sum_{p=1} w_{kp} d_{mkp} + e_m \quad m = 1, \dots, M$$

dove  $w_{kp}$  è un coefficiente che esprime l'importanza attribuita da un generico individuo al fattore  $k$  considerato al livello  $p$ . Tale coefficiente viene detto "utilità parziale" o "part worth". L'insieme dei coefficienti  $w_{kp}$ , con  $p = 1, \dots, P_k$ ,  $k = 1, \dots, K$ , rappresenta quindi il sistema individuale di preferenze per il profilo  $m$ . Le componenti casuali  $e_m$ ,  $m = 1, \dots, M$ , rappresentano le componenti di errore del modello, dovute all'eventuale influenza di fattori noti o non noti non controllati in sede di esperimento, che possono influire sulla preferenza  $Y_m$  espressa dal generico rispondente sul profilo  $m$ . Le componenti casuali  $e_m$  sono poste indipendenti ed identicamente distribuite con media nulla e varianza costante  $\sigma^2$ . Il modello di utilità individuale si configura quindi come un modello di regressione lineare multipla con covariate  $d_{mkp}$  binarie. Si asserisce che nel modello (1) non siano presenti termini di interazione corrispondenti a combinazioni di livelli; è possibile comunque prevedere modelli che comprendono anche le interazioni. Si può esprimere il modello (1) in forma matriciale nel modo seguente:

$$E(\underline{e}) = 0$$

$$\underline{Y} = \underline{Z}\underline{W} + \underline{e} \quad (2)$$

$$\text{Var/ cov}(\underline{e}) = \sigma^2 \text{Im} \quad \text{con Im: matrice identità}$$

Dove  $\underline{Y}$  è il vettore colonna  $m \times 1$  dei punteggi di valutazione osservati per un dato rispondente;  $\underline{Z}$  è la matrice  $m \times P$ , dove  $P = \sum_{k=1}^K P_k$  è il numero totale dei livelli dei fattori, la matrice  $\underline{2}$  è la matrice delle variabili binarie  $d_{mkp}$ ;  $\underline{W}$  è il vettore colonna  $P \times 1$  dei coefficienti incogniti che rappresentano le utilità parziali ed  $\underline{e}$  è il vettore colonna  $m \times 1$  degli errori  $e_m$ .

Tornando all'esempio introdotto precedentemente e considerando i punteggi di preferenza espressi da un dato rispondente il modello (2) sarà del tipo:

$y_1 = 5,2$	=	10	10	10	x	$w_{11}$	+	$e_1$
$y_2 = 7,3$		10	10	01		$w_{12}$		$e_2$
$y_3$		10	01	10		$w_{21}$		$e_3$
$y_4$		10	01	01		$w_{22}$		$e_4$
$y_5$		01	10	10		$w_{31}$		$e_5$
$y_6$		01	10	01		$w_{32}$		$e_6$

$y_7$		01	01	10			$e_7$
$y_8 = 5,7$		01	01	01			$e_8$

La stima dei coefficienti incogniti  $\underline{W}$  può essere ottenuta applicando il metodo dei minimi quadrati ordinari. Risultando però  $Z$  una matrice singolare, per poter ottenere le soluzioni ai minimi quadrati si deve porre una condizione complementare che può consistere ad esempio nel sopprimere, con scelta arbitraria, una colonna in ciascun blocco di variabili binarie della matrice  $Z$ , ponendo di conseguenza pari a 0 il coefficiente corrispondente. Nella matrice  $Z$  viene quindi soppresso un numero di colonne pari al numero di fattori.

Tornando all'esempio precedente il modello diventa:

$y_1 = 5,2$		1	1	1				$e_1$
$y_2 = 7,3$		1	1	0				$e_2$
$y_3$		1	0	1				$e_3$
$y_4$		1	0	0		$w_{11}$		$e_4$
$y_5$	=	0	1	1	x	$w_{21}$	+	$e_5$
$y_6$		0	1	0		$w_{31}$		$e_6$
$y_7$		0	0	1				$e_7$
$y_8 = 5,7$		0	0	0				$e_8$

Verranno quindi stimati i coefficienti  $w_{11}$ ,  $w_{21}$ ,  $w_{31}$ , mentre sono pari a zero i coefficienti  $w_{12}$ ,  $w_{22}$ ,  $w_{32}$ . In generale quindi con tale operazione il modello di regressione diventa:

$$\underline{Y} = Z_0 \underline{W}_0 + \underline{e}$$

dove  $\underline{Y}$  è il vettore colonna  $m \times 1$ ;  $Z_0$  è la nuova matrice non singolare  $m \times (P-K)$ ;  $\underline{W}_0$  è il vettore colonna dei coefficienti  $(P-K) \times 1$  ed  $\underline{e}$  è il vettore colonna degli errori  $m \times 1$ .

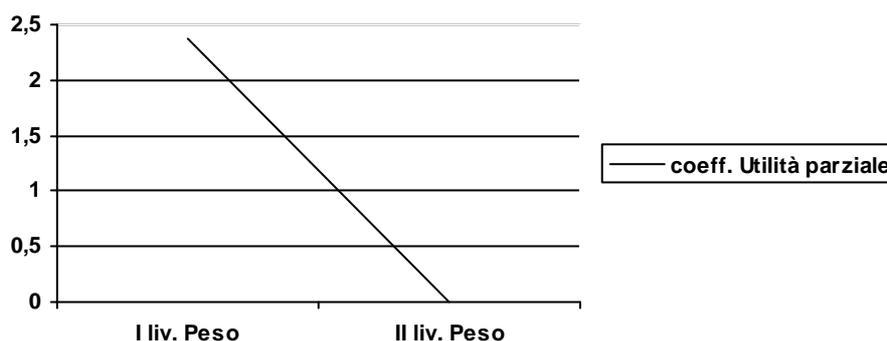
Il vettore delle stime dei coefficienti incogniti secondo il metodo dei minimi quadrati sarà:

$$\underline{W}_0 = (Z_0' Z_0)^{-1} Z_0' \underline{Y}$$

Si osservi che il modello (2) di utilità individuale è stimato per ciascun rispondente. Gli elementi della matrice  $Z_0$  sono fissi per ciascun rispondente mentre varierà il vettore  $\underline{Y}$  dei punteggi di valutazione. Per ciascun rispondente verrà quindi stimato il vettore  $\underline{W}_0$  delle utilità parziali di ciascun livello dei diversi fattori. Tornando all'esempio precedente si suppone che per un dato rispondente siano state stimate le seguenti utilità parziali per il fattore peso:

FATTORE	LIVELLI	COEFFICIENTE DI UTILITA' PARZIALE
PESO	1 Kg	2,37
	2 Kg	0,00

Si osservi che avendo soppresso nella matrice  $Z$  l'ultimo livello dei vari fattori, questo rappresenta il livello di riferimento rispetto al quale sono misurati gli effetti dei restanti livelli. Si possono rappresentare graficamente i coefficienti di utilità parziale stimati per un rispondente nel modo seguente:



Nelle applicazioni di C.A. è inoltre consuetudine valutare l'importanza relativa dei fattori (allo scopo di rendere comparabili le utilità parziali e di pervenire a valori di importanza dei fattori compresi tra 0 e 1) mediante il seguente indice:

$$I_{rk} = \Delta_k / \sum_{k=1}^K \Delta_k \quad k = 1, \dots, K$$

dove  $\Delta_k$  è la differenza (campo di variazione) tra l'utilità del fattore k. L'indice  $I_{rk}$  è tale che  $0 \leq I_{rk} \leq 1$ . L'importanza di un fattore è indicata quindi dalla capacità del fattore di determinare il maggior effetto differenziale nelle utilità parziali.

## 2.4 Analisi grafica dell'influenza degli effetti dei fattori sulla preferenza dei rispondenti

Si è visto precedentemente che il modello di utilità individuale è stimato per ciascun rispondente dando quindi luogo ai vettori  $\underline{W}_0$  delle utilità parziali dei livelli dei vari fattori inclusi nell'esperimento quanti sono i rispondenti. Si suppone che sia d'interesse, in particolare, una valutazione degli effetti dei vari fattori sui giudizi medi espressi dai rispondenti. Si riprende l'esempio introdotto precedentemente e si considera oltre agli effetti principali dei tre fattori A, B, C anche le interazioni fino all'ordine massimo, per cui la matrice del piano fattoriale completo  $2^3$  è del tipo:

PROFILI	FATTORI						
	A	B	C	AB	AC	BC	ABC
1) a b c	+	+	+	+	+	+	+
2) a b	+	+	-	+	-	-	-
3) a c	+	-	+	-	+	-	-
4) a	+	-	-	-	-	+	+
5) b c	-	+	+	-	-	+	-
6) b	-	+	-	-	+	-	+
7) c	-	-	+	+	-	-	+

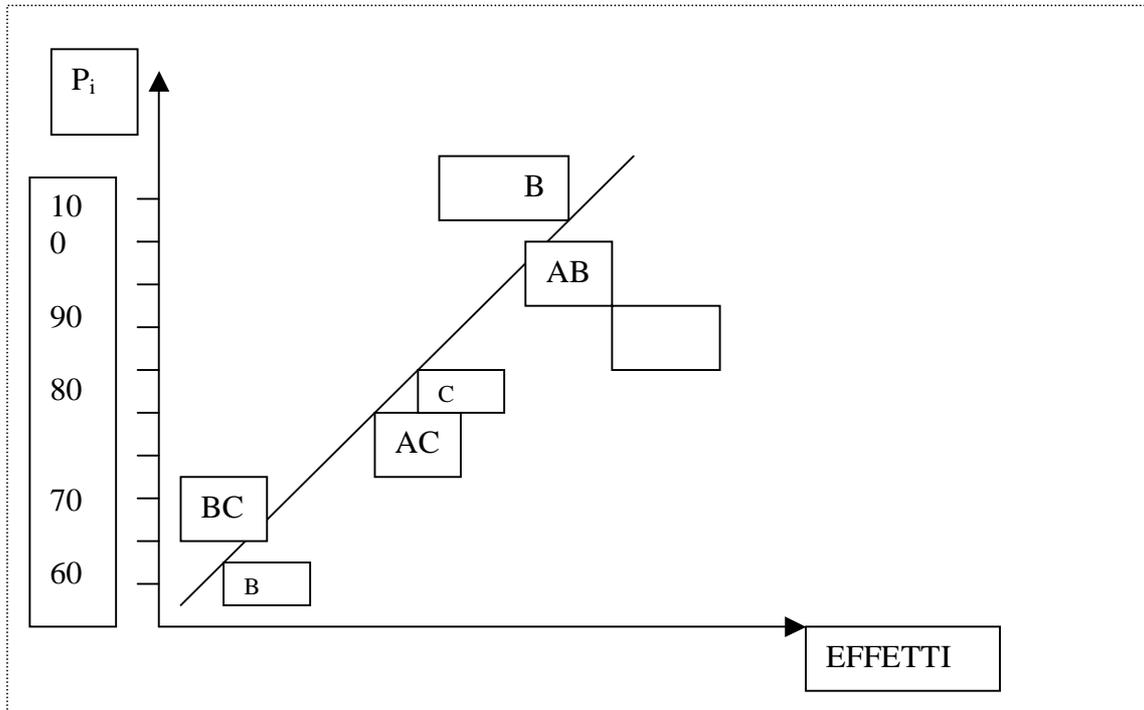
8) (1)	-	-	-	+	+	+	-
--------	---	---	---	---	---	---	---

Si suppone di aver aggregato i rispondenti secondo la media dei punteggi di preferenza rispetto a ciascun profilo di risposta per cui il vettore  $\underline{Y}$  ora contiene le medie di tali punteggi.

$\underline{Y}$	$\bar{y}_1$
	$\bar{y}_2$
	$\bar{y}_3$
	$\bar{y}_4$
	$\bar{y}_5$
	$\bar{y}_6$
	$\bar{y}_7$
	$\bar{y}_8$

Si ipotizza ora di aver stimato le utilità parziali di ciascun livello di ciascun fattore considerando come risposta i punteggi medi dei rispondenti, seguendo la procedura descritta precedentemente. L'interesse a questo punto è di valutare l'influenza dei fattori A, B e C e delle interazioni di tali fattori sui giudizi medi dei profili di prodotto espressi da tutti i rispondenti. Non essendovi gradi di libertà per la stima della varianza dell'errore e dato che la matrice dei profili è non replicata, è possibile evidenziare in modo grafico se un effetto influenzerà o meno la risposta sulla base della sua distanza dalla linea retta ideale tracciata sul NORMAL PROBABILITY PLOT. Dopo aver quindi stimato gli effetti dei fattori e delle interazioni si procede con la costruzione di un grafico probabilistico delle stime degli effetti. Se nessuno influenza la risposta, le stime si comportano come componenti casuali estratte da una distribuzione normale con media nulla, e gli effetti riportati nel grafico giaceranno approssimativamente su di una linea retta. Per disegnare il Normal Probability Plot si ordinano in senso crescente i 7 effetti dei fattori e delle interazioni che rappresentano le ascisse del grafico e si calcolano in corrispondenza i valori delle ordinate dati da:

$$P_i = 100 (i-1/2)/7 \text{ per } i = 1, \dots, 7$$



Dall'ipotetico grafico sopra riportato si può evincere che A è l'unico fattore che sembra influenzare la variabile risposta.

## 2.5 Verifica di ipotesi sugli effetti degli attributi mediante Anova

Si suppone ora che i rispondenti siano stati preventivamente raggruppati in modo da costituire dei gruppi omogenei mediante ad esempio tecniche di CLUSTER ANALYSIS.

In tale situazione è possibile considerare i singoli rispondenti di ogni gruppo omogeneo come replicazioni. Quindi ad ogni singolo profilo di risposta è possibile attribuire una numerosità  $n_m$  di unità sperimentali (soggetti rispondenti) con  $m = 1, \dots, M$ , possibilmente bilanciata, ovvero  $n_m = n$  per ogni profilo. Per esemplificare si considera il caso di un esperimento  $2^2$  replicato 2 volte (due rispondenti). La matrice dei profili e il vettore delle risposte sono i seguenti:

FATTORI			RISPOSTA
A	B	AB	$Y_{MJ}$
+	+	+	$y_{11}$
+	+	+	$y_{12}$
+	-	-	$y_{21}$
+	-	-	$y_{22}$
-	+	-	$y_{31}$
-	+	-	$y_{32}$
-	-	+	$y_{41}$
-	-	+	$y_{42}$

dove  $m$  indicizza il numero dei profili  $m = 1, \dots, 4$  e  $j$  indicizza le repliche,  $j = 1, \dots, n_m$  e nel caso specifico  $j = 1, 2$ .

La corrispondente matrice  $Z$  di variabili binarie per la stima delle utilità parziali dei singoli livelli dei fattori e delle combinazioni di livelli sarà quindi:

$Z =$	1	0	1	0	1	0
	1	0	1	0	1	0
	1	0	0	1	0	1
	1	0	0	1	0	1
	0	1	1	0	0	1
	0	1	1	0	0	1
	0	1	0	1	1	0
	0	1	0	1	1	0
		A		B		AB

Da qui si può procedere secondo la procedura descritta precedentemente per stimare le utilità parziali dei singoli livelli e delle interazioni e per calcolare l'indice  $I_{rk}$  di importanza relativa dei vari fattori.

Nel caso in cui si facciano delle repliche dei profili di prodotto, si può inoltre interpretare le  $y_{mj}$  secondo un modello di ANOVA multifattoriale. Nel caso specifico dell'esempio le  $y_{mj}$  possono essere espresse secondo un modello ANOVA a due vie:

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ij}$$

$$i = 1, \dots, a; \quad l = 1, 2, \dots, b; \quad j = 1, 2, \dots, n$$

dove  $i$  indicizza il numero di livelli del fattore A,  $l$  indicizza il numero di livelli del fattore B e  $j$  il numero di repliche. Mediante l'applicazione dell'analisi della varianza sul modello ANOVA a due vie si può verificare tramite i test F parametrici se gli effetti dei fattori A e B e dalla loro interazione influenzino in modo significativo le risposte sperimentali. L'analisi della varianza per la verifica d'ipotesi sugli effetti dei fattori e dell'interazione è schematizzata nella tabella seguente:

FONTE DI VARIABILITÀ	SOMMA DEI QUADRATI	G.d.l.	MEDIE DEI QUADRATI	TEST F
A	$SS_A$	a-1	$MS_A = SS_A / (a-1)$	$F_{OA} = MS_A / MS_E$
B	$SS_B$	b-1	$MS_B = SS_B / (b-1)$	$F_{OB} = MS_B / MS_E$
AB	$SS_{AB}$	(a-1)(b-1)	$MS_{AB} = SS_{AB} / ((a-1)(b-1))$	$F_{OAB} = MS_{AB} / MS_E$
ERRORE	$SS_E$	ab(n-1)		
TOTALE	$SS_T$	abn-1	$MS_E = SS_E / (ab(n-1))$	

Dove  $SS_A$  è la devianza del fattore A e  $SS_E$  è la devianza dell'errore.

Ad esempio per verificare il seguente sistema di ipotesi:

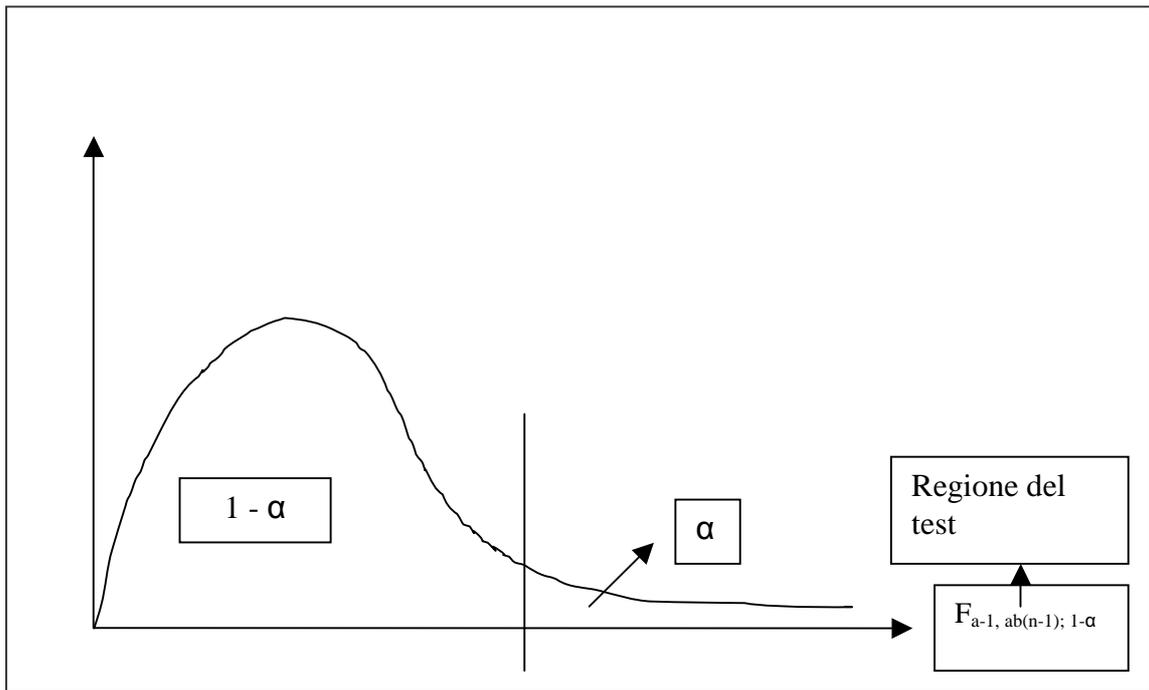
$$M_0 = \alpha_1 = \alpha_2 = \dots = \alpha_a = 0$$

$$M_1 = \alpha_i \neq 0 \text{ per almeno un } i$$

il rapporto  $F_{OA} = M_{SA}/M_{SE}$  si distribuisce sotto l'ipotesi nulla come una  $F_{a-1; ab(n-1)}$  centrale.

Se il valore osservato di  $F_{OA}$  calcolato per la componente osservata  $F_{oss}$  è tale che  $F_{oss} > F_{a-1, (ab(n-1)); 1-\alpha}$

si rifiuterà  $M_0$  al livello di significatività  $\alpha$ . Il grafico seguente mostra l'individuazione della regione critica del test.



Similmente si deriva una verifica d'ipotesi per  $F_{OB}$  e  $F_{OAB}$ .

## 2.6 L'Uso di piani fattoriali nella Conjoint Analysis

Quando il numero dei fattori o il numero dei livelli aumenta, aumenta anche il numero dei profili da sottoporre ai rispondenti. Se ad esempio si hanno cinque fattori su due livelli oppure tre fattori con rispettivamente 2, 3 e 4 livelli, nel primo caso si avrà un numero di profili pari a 32, nel secondo caso 24, rendendo difficoltosa la somministrazione di tutti i profili. Per ovviare a tale inconveniente occorre fare riferimento alle tecniche di riduzione del piano sperimentale, sviluppate nell'ambito

della pianificazione degli esperimenti. In particolare, nel seguito, si considera il caso in cui si voglia condurre uno studio pilota con fattori su due livelli dove il numero dei fattori è pari a 3 o a 4.

Se ad esempio il numero dei fattori è pari a 4, per realizzare un piano fattoriale completo che consenta la stima anche di tutte le interazioni fino a quelle di ordine massimo, bisogna disporre di un numero di profili pari a 16. Un così alto numero di profili può rendere difficile la realizzazione pratica dell'esperimento di C.A., in quanto il rispondente potrebbe fornire giudizi di preferenza per lui significative in tanti profili ed inoltre l'indagine potrebbe essere troppo lunga e stancante per il rispondente.

Si consideri il seguente piano fattoriale  $2^3$ :

PROFILI	EFFETTI FATTORIALI						
	A	B	C	AB	AC	BC	ABC
a b c	+	+	+	+	+	+	+
a b	+	+	-	+	-	-	-
a c	+	-	+	-	+	-	-
a	+	-	-	-	+	+	+
b c	-	+	+	-	-	+	-
b	-	+	-	-	+	-	+
c	-	-	+	+	+	-	+
(1)	-	-	-	+	+	+	-

dove ciascuna prova è indicata con lettera minuscola: se una lettera è presente, allora il corrispondente fattore è fissato al valore più alto se è assente, il fattore è presente al livello basso. La prova con tutti i fattori a livello basso è indicata con (1).

In generale una frazione,  $\frac{1}{2}$  di un piano  $2^k$  contiene  $2^{k-1}$  prove e viene chiamato piano fattoriale frazionario  $2^{k-1}$ . Ad esempio si considera il piano  $2^{3-1}$ , cioè la frazione  $\frac{1}{2}$  del piano  $2^3$ . Si suppone di scegliere le quattro prove a, b, c, abc come frazione  $\frac{1}{2}$  del piano  $2^3$ . Si selezionano le prove che forniscono un segno positivo per l'effetto ABC. Dunque ABC è detto generatore di questa particolare frazione. Considerando  $I = ABC$  relazione di definizione del piano ed indicando in generale con  $I_A$  la stima dell'effetto del fattore

A, si può dimostrare che  $I_A = I_{BC}$ ,  $I_B = I_{AC}$ ,  $I_C = I_{AB}$ . Conseguentemente non si può differenziare tra A e BC, B e AC e C e AB. Due o più effetti che hanno questa proprietà si dicono “alias”. Nel piano considerato  $2^{3-1}$ , A e BC, B e AC, C e AB sono alias.

In molte situazioni pratiche, sarà possibile scegliere la frazione in modo tale che gli effetti principali e le interazioni di basso ordine siano alias d’interazioni di ordine alto (che saranno presumibilmente trascurabili). La struttura degli alias di questo piano, si determina a partire dalla relazione di definizione  $I = ABC$ . Moltiplicando ogni effetto per la relazione di definizione, si determinano gli alias di quell’effetto. Ad esempio l’alias di A è:

$$A = A \times ABC = A^2BC = BC$$

Si noti che  $A^2 = I$  e quindi  $A^2BC$  restituisce la colonna BC. Si consideri ora una frazione  $\frac{1}{2}$  del piano  $2^4$  e si supponga di utilizzare il piano  $2^{4-1}$  con  $I = ABCD$  per analizzare i 4 fattori A, B, C, D. In questo piano gli effetti principali sono alias di interazioni fra i tre fattori:

$$A \times I = A \times ABCD$$

$$A = A^2BCD$$

$$A = BCD$$

Analogamente:

$$B = ACD$$

$$C = ABD$$

$$D = ABC$$

Le interazioni fra due fattori sono alias fra di loro:

$$AB = CD$$

$$AC = BD$$

$$AD = BC$$

Quindi in questo piano i fattori principali sono stimabili in modo non distorto se si assume che l’effetto delle interazioni di ordine tre è trascurabile, mentre gli effetti delle interazioni di ordine due non saranno stimabili in modo non distorto poichè le interazioni sono confuse a coppie.

Supponendo quindi di dover condurre un esperimento di C.A. con quattro fattori principali su due livelli, se è possibile ipotizzare che l'effetto delle interazioni di ordine superiore o uguale al terzo sia trascurabile, si possono ottenere le stime delle utilità parziali dei livelli dei fattori, applicando un piano frazionato  $2^{4-1}$  con relazione definente  $I = ABCD$ . Questo permette quindi di sottoporre a ciascun rispondente un numero di profili pari ad 8 in luogo dei 16 che necessiterebbe un piano completo  $2^4$ . Si noti che tutti i piani frazionati o completi finora considerati hanno l'importante proprietà di essere piani di tipo ortogonale, ovvero sia  $X$  la matrice dei profili, si ha che  $(X^T X)^{-1} = 2^k I_k$  dove  $I_k$  è la matrice identità di ordine  $k$ , con  $k$  numero di attributi. L'utilizzo di piani completi o frazionati ortogonali per condurre una Full Profile C.A. permette di rendere minima la varianza degli stimatori ai minimi quadrati utilizzati per la stima delle utilità parziali. Sia  $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$  il modello lineare di riferimento, con  $E(\underline{\varepsilon}) = \underline{0}$  e  $\text{Var}(\underline{\varepsilon}) = \sigma^2 I_m$ , dove  $m$  è il numero dei profili, si ha che  $\text{Var}(\underline{\hat{\beta}}) = \sigma^2 (X^T X)^{-1}$  dove  $\underline{\hat{\beta}}$  è la stima ai minimi quadrati di  $\underline{\beta}$ . Nella formula della matrice di varianza e covarianza dei  $\underline{\hat{\beta}}$  compaiono due quantità  $\sigma^2$  e  $(X^T X)^{-1}$ . La soluzione di minimo di tale quantità è realizzata solo nel caso in cui la matrice dei profili sia ortogonale.

Per concludere, nell'adattare una Full Profile C.A. diventa cruciale la scelta della matrice dei profili che nel caso in cui il numero di attributi sia relativamente elevato, dovrebbe essere scelta secondo un piano frazionato di tipo ortogonale.

## 2.7 Analisi bibliografica

La ricerca bibliografica di articoli riguardanti la Conjoint Analysis ha portato a risultati interessanti per capire quali siano stati i metodi più utilizzati nei vari ambiti di ricerca nel corso degli anni. Infatti i documenti (più di un centinaio) sono stati suddivisi per argomento, per metodo applicativo e per metodo di somministrazione del questionario.

La tabella 2.1 sotto riportata riassume il numero di articoli divisi per metodologia usata rispetto ai vari indirizzi di applicazione: si può notare come in tutti i vari settori venga utilizzata la full profile come prova del fatto che quest'ultima tecnica sia ritenuta quella che garantisce risultati più attendibili. Infatti si osserva che l'adaptive conjoint analysis non solo è meno usata della full profile ma anche non si estende a tutti i settori

analizzati. L'unica tecnica che si avvicina come utilizzo alla full profile è la conjoint based conjoint analysis con ben 26 articoli che trattano dell'utilizzo di quest'ultima.

METODOLOGIA TIPO APPLICAZIONE	FULL PROFILE	TRADE OFF MATRIX	ADAPTIVE CONJOINT ANALYSIS	HYBRID CONJOINT ANALYSIS	CONJOINT BASED CONJOINT ANALYSIS	ALTRO	NON CLASSIFICABILE, NON PRECISATA	TOTALE
SERVIZI	15	1	2	1	4	3	5	31
BANCA	7				1		2	10
MEDICO SANITARIO	3		1		5		1	10
TURISMO	6				1		1	8
AGRICOLTURA/ ALLEVAMENTI	2		1					3
INFORMATICA	4				1			5
ELETTRONICA DI CONSUMO	1			1		1	1	4
MERCATO DEL LAVORO, OCCUPAZIONE	4		1	2			2	9
AUTOMOBILISTICO	1			1		1	3	6
TRASPORTI E LOGISTICA	2				1			3
FOOD & DRINK	12		1		4		3	20
TELECOMUNICAZIONI	2					1		3
ALTRI SETTORI	10	1	1		8	5		25
ECOLOGICO/AMBIENTALE	1				1		1	3
<b>TOTALE</b>	<b>70</b>	<b>2</b>	<b>7</b>	<b>5</b>	<b>26</b>	<b>11</b>	<b>19</b>	<b>140</b>

**Tabella 2.1**

Facendo un'analisi accurata della tabella riga per riga si può notare come la Conjoint Analysis sia più utilizzata nei settori come: servizi, banca, medico-sanitario, mercato del lavoro e food & drink, settori molto diversi tra loro. Questo dimostra come la C.A. si sia sviluppata e divulgata negli anni ricoprendo un ruolo sempre più importante nelle ricerche di mercato.

Un altro aspetto molto interessante è quello di studiare i diversi intrecci tra le metodologie di Conjoint Analysis e i vari metodi di somministrazione del questionario rappresentati nella tabella 2.2. Anche qui come nel grafico precedente la full profile è l'unica tecnica che viene utilizzata in tutti i vari tipi di indagine nonostante sia la più laboriosa e onerosa tra le tecniche. La tabella evidenzia inoltre che in corrispondenza dell'adaptive conjoint analysis viene usato il computer interactive interview come unico metodo di somministrazione del questionario. Oltre alla full profile ed all'adaptive un'altra tecnica molto utilizzata è la choice based conjoint analysis.

METODOLOGIA TIPO DI CONTATTO E DI INTERVISTA	FULL PROFILE	TRADE OFF MATRIX	ADAPTIVE CONJOINT ANALYSIS	HYBRID CONJOINT ANALYSIS	CHOICE BASED CONJOINT ANALYSIS	ALTRO	NON CLASSIFICABILE/ NON PRECISATA	TOTALE
PERSONAL INTERVIEW	21	2			7	2	5	37
QUESTIONARIO DISTRIBUITO	11			1	2		1	15
QUESTIONARIO INVIATO PER POSTA	10				6		4	20
COMPUTER INTERACTIVE INTERVIEW	1		7		4	3		15
TELEPHONE-MAIL-TELEPHONE	2			1		1		4
ALTRO/METODI VARI	5			2			1	8
NON INDICATO	20			1	7	5	8	41
TOTALE	70	2	7	5	26	11	19	140

**Tabella 2.2**

La tabella mette in luce l'utilizzo maggiore della tecnica personal interview, infatti per ottenere un buon risultato dall'applicazione della C.A. i dati devono essere chiari e

corretti; quindi intervistando di persona i vari soggetti si riduce la probabilità di avere errori o equivoci nelle risposte. Purtroppo questo metodo richiede molto tempo sia per l'intervistato che per il ricercatore ed inoltre è alquanto costoso, ecco perché per ovviare a questo problema nel corso degli anni si sono sviluppate delle tecniche alternative meno efficaci in qualità come: il questionario distribuito, inviato per posta o telefonico e il computer interactive interview.

## **2.8 Alcuni problemi sull'aggregazione dei rispondenti nella Conjoint Analysis**

Per costruire una classifica globale partendo dalle n classifiche parziali ottenute dai rispondenti, l'unico metodo oggi usato in letteratura è la media aritmetica. Si può affermare questo anche grazie alle opinioni date in merito da diversi prof.ri esperti in materia come: prof. Luigi Burigana (docente di Psicometria Corso Avanzato all'università di Padova), prof. Amedeo De Luca (docente all'università Cattolica del S.Cuore di Milano), prof. Carlo Natale Lauro (docente di Statistica all'università di Napoli), prof. Giulio Vidotto (docente di Psicometria all'università di Padova), ad alcuni articoli [Moore e William L. 1980] e libri di testo [Grandinetti, 2002 e De Luca, 2000]. Infatti il prof. De Luca nel suo libro afferma che: “ per definire un modello aggregato di utilità, idoneo a descrivere la struttura di preferenza comune di un gruppo di rispondenti, con l'approccio tradizionale si considera semplicemente la media aritmetica dei coefficienti di utilità individuale stimati [Hagerty, 1985].

Anche se questo metodo è largamente il più usato depaupera e distorce i risultati dell'analisi perché l'utilità media, in caso di gruppi non omogenei, non rispecchia fedelmente i singoli giudizi degli individui. Infatti Moore e William L. nel loro articolo affermano che: *”At the other extreme of the aggregation continuum is the case in which the preferences ratings are pooled across all respondents and one overall utility function is estimated. These pooled utility estimates are the same as the average of the individual utility estimates. The results of this type of analysis are easy to explain(e.g., on average, travel time is much more important than trip cost in determining a transportation mode). A potential problem with pooled analysis is termed the ‘majority*

*fallacy' [Kuehn and Day, 1962]. It occurs when the item chosen by the average customer is not the one chosen most often. The majority fallacy is caused by heterogeneity of preferences; for example, if half of the people like large cars and the other half like small cars, the average person may like medium-sized cars best, even though no real person wants one. This problem has been demonstrated in a conjoint analysis setting by Huber and Moore (1979). As expected, heterogeneity of preference reduces the predictive power of the aggregate model; for example, Wittink and Montgomery (1979) found that a pooled model could predict only 46% of the students' job choices (compared with 63% with individual models)."*

Nel quarto capitolo si presenterà un metodo statistico non parametrico di aggregazione (NPC ranking) come alternativa alla media aritmetica.



## **Capitolo 3**

### **APPLICAZIONI DELLA CONJOINT ANALYSIS NEL SETTORE BANCARIO**

#### **3.1 Introduzione**

La Conjoint Analysis (C.A.) è una tecnica che è stata creata allo scopo di analizzare le preferenze dei possibili consumatori, su varie alternative di prodotto. Le sue notevoli caratteristiche di flessibilità e adattabilità hanno fatto sì che, in questi ultimi anni, si siano compiute numerose indagini su servizi finanziari con questa metodologia. In particolare gli studi eseguiti su svariati aspetti del settore bancario e finanziario sono stati molteplici. In generale si possono suddividere i vari studi per argomento in tre grandi aree: la prima legata ad indagini inerenti alle carte di credito, la seconda area legata agli investimenti finanziari; la terza, generica in cui far confluire vari argomenti legati all'ambito bancario, come l'analisi sulla percezione, da parte del cliente, del servizio percepito o sui conti correnti.

Si vuole riportare in questo capitolo un compendio delle pubblicazioni più interessanti apparse nelle riviste del settore bancario e di marketing. Tutto questo per fornire una panoramica degli studi fatti ed indicare le fonti dalle quali si sono prese le indicazioni per l'analisi.

In questo capitolo si esaminerà l'evoluzione degli studi riguardanti gli investimenti finanziari e la banca in generale. Di ogni articolo si presenterà una prima parte introduttiva, un cenno sulla metodologia e al campione dell'indagine. Inoltre sono riportate le tabelle più significative con i risultati e alla fine è presente una sintesi delle conclusioni tratte dagli autori degli articoli.

### **3.2 Gli investimenti finanziari**

In questo paragrafo si sono considerati gli articoli riguardanti gli investimenti finanziari, riferiti sia al singolo consumatore sia alle società di capitali. Le varie indagini hanno analizzato i criteri di scelta degli investimenti finanziari cercando di capire quali fossero le caratteristiche che maggiormente attirassero un investitore (singola persona o società di capitali).

Nel primo articolo che risale al 1996 [Muzyka, Birley e Leleux, 1996] si cerca di capire quali siano i fattori che guidano le scelte di investimento delle società di Venture Capital, società che forniscono fondi alle imprese nascenti. L'obiettivo è di valutare i fattori decisionali e la somiglianza tra le varie società presenti in Europa.

La metodologia usata è stata un'indagine di Conjoint Analysis full profile; la lista di attributi e livelli si è ottenuta eseguendo interviste con dirigenti di società e analizzando la letteratura precedente. In questo modo si è costruita una lista di 35 attributi raggruppati per chiarezza espositiva in sette insiemi relativi a criteri: finanziari, di prodotto-mercato, strategici-competitivi, di fondo, di gestione, di competenza nella gestione e criteri sulla natura dell'investimento.

Ai rispondenti sono stati richiesti 53 confronti e, una prima fase pilota, ha permesso di modificare qualche piccolo particolare legato più che altro alla terminologia. Con le migliorie apportate, il test è stato somministrato in vari paesi a diversi manager ottenendo 73 questionari completi.

Dall'indagine è risultato che i cinque criteri del Management Team sono tra i principali fattori decisivi, la Leadership ability dell'imprenditore e del team di gestione sono i primi due. Oltre a questi è stato decisivo che il prodotto della società da finanziare riuscisse a mantenere la posizione nel mercato e che il management fosse capace. Inoltre sono risultati come significativi anche i fattori finanziari legati alla capacità di produrre reddito e di ripagare il finanziamento.

In sintesi è risultato che le società di capitali valutano tutte le opportunità di buoni affari senza avere pregiudizi verso settori poco presenti nel loro portafoglio oppure legati alle dimensioni o al settore del business.

L'importanza maggiore si è data al management, alla situazione finanziaria e alla tipologia di prodotto.

Nella tabella 3.1 sono riportati gli attributi, i livelli e nella prima colonna l'ordinamento finale in base all'importanza. L'analisi delle risposte ha fornito informazioni sufficienti a definire tre tipologie di società:

- a) Investitori nazionali che preferiscono le aziende del proprio paese potendo così meglio interpretare la realtà economica e la concorrenza.
- b) Un insieme di società, molto sensibili alla tipologia del business, che puntano a settori ben precisi e non sconfinano in altri a loro poco noti.
- c) La maggioranza delle aziende che valutano nel complesso i possibili affari usando parametri comuni e abbastanza ovvi come: la quota di mercato, la capacità della società e la profittabilità

Ord.	Financial criteria	Trade-off Options		
12	Time to breakeven	<1,5 years	1.5 to 3 years	> 3 years
20	Time to payback	<3 years	3 to 5 years	> 5 years
11	Expected rate of return	<16%	16% to 25%	>25%
9	Ability to cash out	Strong	Moderate	Weak
Product-Market criteria				
19	Degree of market already established	Exists	Marginal	New
29	Market size	Large	Medium	Small
33	Seasonality of product-market	Sensitive	Somewhat	Insensitive
30	Sensitivity to economic cycle	Sensitive	Somewhat	Insensitive
18	Market growth and attractiveness	<15%	16 to 25%	>25%
17	Uniqueness of product and technology	High	Medium	Low
27	National location of business	Same	Adjacent	Immaterial
10	Degree of product-market understanding	High	Medium	Low
Strategic-Competitive criteria				
24	Ease of market entry	Easy	Moderate	Difficult
14	Ability to create postentry barriers	Strong	Moderate	Weak
6	Sustained share competitive position	Strong	Moderate	Weak
26	Nature and degree of competition	Strong	Moderate	Weak
25	Strength of suppliers and distributors	Strong/clear	Moderate	Unclear
Fund criteria				
15	Business meets fund constraints	Match	Partial match	Mismatch
28	Business and product fit with fund portfolio	Strong	Moderate	Weak
24	Ability of investors to influence nature of the business	Strong	Moderate	Weak
35	Location of business relative to fund	Close	Within reach	Immaterial
Management team criteria				
2	Leadership potential of management team	Strong	Moderate	Weak/uncertain
1	Leadership potential of lead entrepreneur	Strong	Moderate	Weak/uncertain
3	Recognized industry expertise in management team	High	Moderate	Low/uncertain
4	Track record of lead entrepreneur	Strong	Moderate	Weak/uncertain
5	Track record of management team	Strong	Moderate	Weak/uncertain
Management competence criteria				
7	Marketing/Sales capability of team	Strong	Moderate	Weak
16	Process/Production capability of team	Strong	Moderate	Weak
8	Organizational/Administrative capability of team	Strong	Moderate	Weak
13	Financial/Accounting capability of team	Strong	Moderate	Weak
Deal criteria				
23	Stage of investment required	Late	Middle	Early
32	Number and nature of co-investors in the deal	Many/Complex	Few	None/simple
31	Ability to syndicate the deal	Strong	Moderate	Low
34	Scale and chance of later rounds of financing	Significant	Moderate/likely	Small/unlikely
22	Importance of unclear assumptions	High	Medium	Low

**Tabella 3.4 Criteri di analisi con i rispettivi livelli e l'ordine di importanza ottenuto.**

Da queste analisi si possono trarre alcune conclusioni sulle società di capitale europeo: innanzi tutto che non esistono differenze tra stati e che la realtà europea è omogenea, inoltre le società sono più orientate all'area finanziaria di quelle americane.

Altre considerazioni possono essere d'aiuto agli imprenditori che cercano capitali come creare un buon team di management ed evidenziare bene le proprie capacità e opportunità di mercato, preoccupandosi un po' meno degli aspetti strettamente finanziari. Inoltre è importante non fermarsi nella ricerca di capitali nel proprio paese

ma cercare anche all'estero, poiché le società di capitali non vedono come grossi ostacoli le differenze di legislazione o le differenti pratiche finanziarie tra paesi.

Nel secondo articolo trattato, l'argomento rientra un po' forzatamente nell'ambito degli investimenti perché parla di mutui per acquistare casa e risale al 1998 [Malaga e Buch, 1998].

La panoramica introduttiva spiega come comunemente le persone spendano molto tempo a cercare la casa ideale, valutando moltissimi fattori: dalla posizione vicino alle scuole, al quartiere, alle condizioni delle fondamenta e via dicendo, mentre è dedicato poco tempo a scegliere il giusto mutuo. L'indagine vuole capire come i clienti scelgano quest'ultimo. Le variabili dell'esperimento sono state ricavate dalla letteratura sull'argomento e dagli operatori del settore. Gli attributi e i livelli sono riportati in tabella 3.2 con le valutazioni dei rispondenti senza un'ipoteca e nelle successive colonne di quelli che ne hanno già stipulata una, nel senso che hanno già avuto un'esperienza con il soggetto dell'indagine.

Variable	Gruppo senza esperienza		Gruppo con esperienza	
	Utility	Import.	Utility	Import.
Type of mortgage				
Fixed	1,334722222		0,3418	
Variable	-0.1922	0,725	-0.3418	21,9800
Number of years in mortgage				
15 years	-0,0224		-0.0448	
30 years	0,0224	1,96	0,0448	3,1400
Number of points				
No points	0,272		0,4245	
Two points	0,0102		0,0443	
Three points	-0,2822	24,28	-0.4688	28,7100
Additional costs				
No additional costs	0,4705		0,5033	
\$1-500 additional	0,23		0,1829	
\$501-1,500 additional	-0,2653		-0.2493	
\$1,501+ additional	-0,4351	39,67	-0.4368	30,2200
Knowledge of lender				
Know/heard of lender	0,1969		0,2480	
Don't know of lender	-0,169	17,25	-0.2480	15,9500
Constant	2,6266		2,4714	

**Tabella 3.5** Attributi e livelli con importanza e utilità

Il tasso d'interesse non è stato incluso nelle variabili perché il periodo di ricerca di un'ipoteca è relativamente breve e, per questo, esso non varia di molto. Inoltre tutti vorrebbero il mutuo al minor tasso possibile, ma l'indagine vuole valutare altri fattori perciò questa variabile andrebbe a distorcere in maniera marcata le risposte.

La metodologia usata è una Conjoint Analysis full profile con un piano ortogonale frazionato che ha ridotto da 96 a 16 i profili da somministrare, il campione è stato composto di 158 studenti di un master MBA, considerandoli sufficientemente informati sull'argomento ipoteche, di questi, 98 avevano già stipulato un'ipoteca. I risultati evidenziano innanzi tutto l'utilità diversa tra i due gruppi, si nota come chi non ha esperienza di ipoteche preferisca l'ipoteca ideale (fissa, di trent'anni, senza commissione, senza costi aggiuntivi, con conoscenza dell'erogatore dell'ipoteca) mentre chi ha già esperienza assegni pesi diversi, conoscendo la realtà del mercato. Queste considerazioni offrono spunti di riflessioni su come le ipoteche siano valutate dai due gruppi, la variabile più desiderata è l'assenza di spese aggiuntive, la seconda è i "points" cioè quanti punti percentuale bisogna pagare rispetto all'ammontare dell'ipoteca a chi fornisce il prestito.

Le considerazioni conclusive indicano come, sempre considerando che il tasso d'interesse risulti una forte arma competitiva, società conosciute o collegate a società immobiliari famose possano sfruttare la loro fama facendo pagare un po' di più. Inoltre può essere utile creare due tipi di offerte: per chi non ha mai stipulato un'ipoteca e per chi ha già avuto esperienza.

Nel 1999 è stato pubblicato un articolo inerente alle strategie delle società di capitali [Shepherd, 1999]. Il lavoro parte dall'analisi della letteratura per individuare le strategie usate dalle società di capitali per valutare i nuovi business, queste ipotesi sono poi analizzate con un'analisi di conjoint e infine commentate. Le ipotesi create sono otto e sono:

1. La possibilità di sopravvivere nel mercato è maggiore per una nuova impresa se si basa su fattori di successo stabili.
2. La probabilità di sopravvivere nel mercato cresce se si è inseguitori piuttosto che pionieri in un nuovo settore.
3. Si hanno più possibilità con tempi di sviluppo lunghi piuttosto che sul breve termine.

4. Si sopravvive più facilmente in settori dove la concorrenza è minore.
5. La probabilità di restare nel mercato cresce se si ha un livello di conoscenze elevato.
6. Hanno più possibilità chi possiede alte competenze industriali.
7. Le società di capitali valutano con più attenzione le competenze industriali che i fattori di successo o il momento d'entrata nel mercato.
8. Le società di capitali valutano molto le conoscenze possedute rispetto al resto.

Basandosi su queste considerazioni, è stato creato un questionario somministrato tramite mail e di persona, ai dirigenti di 65 società di capitale australiane, particolarmente attive nell'investire in Seed o Start-up. I due metodi di somministrazione non hanno generato differenze nelle valutazioni e sono stati aggregati. Dalle ipotesi si sono decisi gli attributi e i livelli e si sono aggiunti due attributi come variabili di controllo (Mimicry-imitazione, Scope-scopo).

La valutazione da dare era su quante probabilità avesse un dato profilo di una nuova azienda di sopravvivere per dieci anni nel mercato. Ogni intervista consisteva di 32 profili ottenuti con un piano ortogonale frazionato più altri sette hold-outs.

Criteria	Levels
Key Success Factor Stability	High Low
Timing of Entry	Pioneer Late Follower
Lead Time	Long Short
Competitive Rivalry	High Low
Educational Capability	High Low
Industry Related Competence	High Low
Control Variables Scope	Broad Narrow
Entry Wdge Mimicry	High Low

**Tabella 3.3**

Definition
requirements necessary for succes will not change radically during industry development
requirements necessary for succes will change radically during industry development
enter a new industry first
enter a industry late in the industry's stage of development
an extended period of monopoly for the first entrant prior to competitors entering this industry
a minimal period of monopoly for the first entrant prior to competitors entering this industry
intense competition among industry members during industry development
little competition among industry members during industry development
considerable resources and skills available to overcome market ignorance through education
few resources or skills available to overcome market ignorance through education
venturer has considerable experience and knowledge with the industry being entered on a related industry
venturer has minimal experience and knowledge with the industry being entered on a related industry
a firm that spreads its resources across a wide spectrum of market, e.g. many segments of the market.
a firm that concentrates on intensively exploiting a small segment of the market,e.g. targeting niche.
considerable imitation of the mechanisms used by other firms to enter this, or any other industry, e.g. a franchisee.
minimal imitation of the mechanisms used by other firms to enter this, or any other industry, e.g. a introducing a new product.

Decision Criteria	B Coefficient	Z Score	Omega Squared
Key Success Facror Stability	0,88	10,52*	0,05
Timing of Entry	0,69	3,57*	0,07
Lead Time	0,6600	8,02*	0,0300
Competitive Rivalry	0,9700	12,14*	0,0600
Educational Capability	1,1500	14,04*	0,1000
Industry Related Competence	2,1400	37,77*	0,2900
Control Variables			
Scope	0,0400	0,1800	0,2000
Mimicry	0,2300	2,16*	0,2000
		Significant p<0,05	

**Tabella 3.4**

Nella tabella 3.4 sono riportati i dati risultanti, Z score indica che i valori superiori a 1.65 sono significativi nelle considerazioni dei manager. Il valore Omega Squared indica l'importanza assoluta che grazie ai valori conferma le ipotesi 7 e 8 dando il peso più rilevante alle considerazioni sulle competenze industriali e sulle conoscenze del management della nuova società.

Le conclusioni che si possono trarre, indicano che c'è molta somiglianza tra il metodo di giudizio con cui le società di capitali giudicano le aziende nascenti e le strategie presenti nella letteratura, attribuendo ad entrambe una gestione competente e capace di maggiori sicurezze.

Il lavoro più recente è del 2003 ed analizza le preferenze individuali degli investitori australiani [Clark-Murphy e Soutar, 2003]. In questo caso i ricercatori analizzano la realtà australiana, in cui è largamente diffusa la pratica dell'investire in borsa, per capire quali siano i fattori che guidano le scelte degli investitori. L'Australia è la nazione principale come numero di azionisti, prima ancora dell'Inghilterra e dell'America. Poche sono le ricerche svolte per comprendere come le persone operino le loro scelte, ma l'argomento è d'interesse crescente perché aumentano le possibilità d'investimenti e di conseguenza si alza la richiesta d'informazioni. Nel contesto australiano fra l'altro non esiste un sistema pensionistico pubblico ed è responsabilità del singolo crearsi un capitale su cui poter disporre alla fine del periodo lavorativo, perciò sempre più importanti sono le ricerche volte a capire cosa gli investitori cerchino e desiderino.

Le teorie attuali indicano che l'investitore sceglie con completa razionalità l'investimento, basandosi sulla sua propensione al rischio e sul ritorno atteso, ma come ben si può immaginare le persone non sono mai totalmente razionali e nella scelta c'è sempre un fattore soggettivo.

L'analisi è stata condotta con un'indagine ACA (Adaptive Conjoint Analysis) perché meglio si adattava all'alto numero di attributi da valutare. I rispondenti sono stati investitori individuali caratterizzati da un forte interesse personale per i temi finanziari, nonché attivi nella gestione dei loro investimenti. Queste persone sono state scelte tra i soci dell'ASA (Australian Shareholders' Association) un'associazione senza fini di lucro che offre informazioni agli investitori. Sono stati distribuiti 488 questionari e sono risultati validi 361.

Attributi e livelli dell'indagine sono stati ricavati dalla letteratura e comprendono sia parametri prettamente finanziari e sia legati alle conoscenze dell'investitore o a chi ha consigliato una particolare azione. Questi attributi sono stati testati su un gruppo di studenti di un MBA per capire le difficoltà di comprensione. L'elemento più rilevante è stata l'enorme variazione dei dati al variare delle condizioni di mercato, questo ha suggerito di porre attenzione al periodo in cui il test è stato svolto.

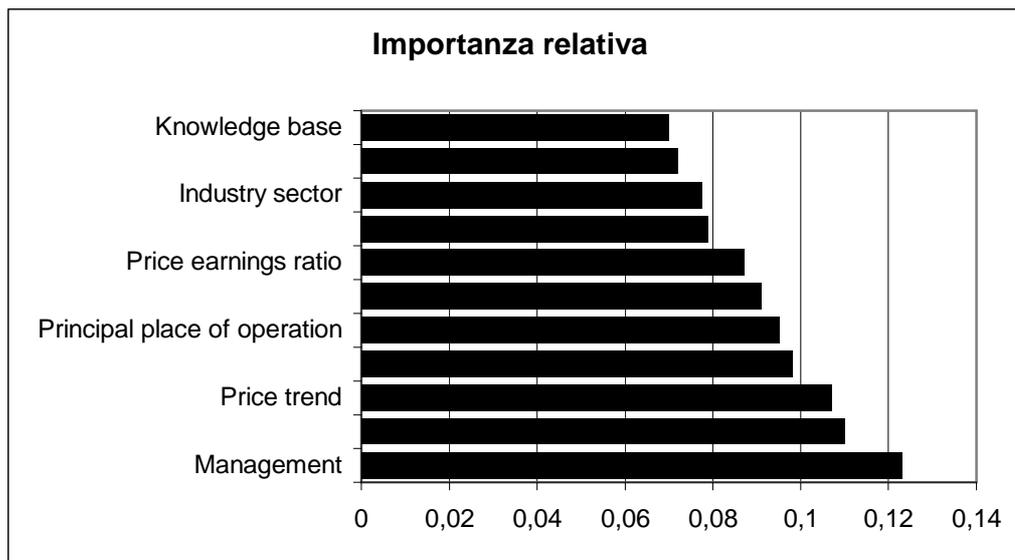
Nella tabella 3.5 sono riportati attributi, livelli e utilità calcolati.

I dati ottenuti provengono per lo più da uomini di età superiore ai 45 anni con interessi di lungo termine e che investono in borsa da molti anni. La maggioranza di questi ricerca nell'investimento un compromesso tra reddito e crescita del capitale.

I risultati ottenuti (grafico 3.1) segnalano che il Management è la variabile più importante, questo conferma il risultato di altre indagini in cui la qualità dei dirigenti è considerata fondamentale anche dagli analisti finanziari. Al secondo posto troviamo la situazione di mercato, spesso legata alla qualità della direzione. Al terzo posto c'è l'andamento dell'azione, il che è abbastanza sorprendente perché altri fattori più strettamente finanziari sono valutati meno importanti.

No.	Name	No.	Text	Utility
1	Dividend	1	Pays a regular dividend	110,5
		2	Not currently paying a dividend	1,42
2	Yield	1	High yield	101,5
		2	Average yield	62,31
		3	Low yield	2,67
3	Management	1	Management has a track record in this industry	106,6
		2	Some managers have a high public profile	19,91
		3	Management are known to you personally	33,29
4	Industry sector	1	Sector has a history of slow but steady growth	65,74
		2	Sector is subject to periods of growth and contraction	22,28
		3	Stock is in an emerging industry sector	58,13
5	Knowledge base	1	You have personal knowledge or experience of the industry	51,76
		2	You use the company's products	16,46
		3	You know of the company's products	31,7
6	Price trend	1	Price has been rising in the last few months	64,61
		2	Price has been falling in the last few months	22,65
		3	Price has been steady in the last few months	44,46
7	Price earnings ratio	1	High price earnings ratio	25,4
		2	Average price earnings ratio	58,06
		3	Low price earnings ratio	62,91
8	Price volatility	1	Price is often volatile	23,4
		2	Price is usually steady	54,35
9	Market status	1	Blue chip stock	98,47
		2	Growth stock	107,4
		3	Speculative stock	8,36
10	Source of recommendation	1	Recommended by your stockbroker	91,02
		2	Recommended in the financial press	80,5
		3	A tip or rumour from a friend	6,76
11	Principal place of operation	1	Company mainly operates in Australia	90,11
		2	Company mainly operates in Asia	10,93
		3	Company mainly operates in Europe or the USA	49,75

**Tabella 3.5: tabella delle utilità**



**Grafico 3.1 sull'importanza relativa**

Risulta che gli investitori si informino bene sulle aziende e spesso investano su aziende che conoscono o di cui usano i prodotti.

In particolare è preferita un'azienda di cui si conosce personalmente il management; a livello di tipologia di società si preferiscono quelle quotate nei listini migliori (Blue Cips). A livello di prezzo viene preferita l'azione con un valore in crescita. La fonte preferita d'informazioni è il promotore finanziario e la piazza finanziaria preferita è quella australiana. Questo è dovuto alla crisi asiatica degli anni 1997-98 di cui c'è ancora fresco ricordo.

Il fatto che un'azienda fornisca dividendi regolari è considerato importante nelle scelte, la preferenza per un basso rapporto prezzo-guadagno, segnala come i rispondenti siano attenti a valutare le opportunità del mercato. Con "Yield" viene inteso l'insieme di rendimento e crescita del capitale, e ovviamente le preferenze vanno per un alto rendimento complessivo.

A livello di settore sono preferite quelle realtà in cui la crescita è moderata ma stabile, così anche a livello di prezzo delle azioni sono preferite quelle più stabili e con bassa volatilità. Le aziende preferite sono quelle di cui si ha una qualche conoscenza diretta o indiretta tramite i prodotti.

In conclusione si può affermare che i parametri più importanti non sono quelli strettamente finanziari come la percezione di qualità del management, il tipo di azione e

il recente andamento del valore delle azioni. Gli intervistati sono propensi ad accettare un certo margine di rischio ma non sono propensi ad usare strumenti speculativi.

### **3.3 Le banche**

In questo paragrafo si sono raccolti articoli di diversi argomenti, senza un preciso filo conduttore, se non l'uso della Conjoint Analysis per studiare le preferenze dei clienti.

La prima pubblicazione analizzata è del 1996, tratta di un'indagine per creare un nuovo servizio finanziario [Gómez, 1996]. Lo studio vuole illustrare l'utilità della tecnica di conjoint per creare nuovi servizi. Nel caso in esame il servizio è indirizzato ai rivenditori ed è volto a valutare l'uso di un lettore di carte clienti, di smart card, contenenti il profilo del cliente e di un sistema computerizzato di accreditamento alla banca che offre il sistema.

La banca si propone di offrire una linea di credito in tempo reale al cliente che ha bisogno di un finanziamento nel momento dell'acquisto, basandosi sui dati personali del cliente presenti nella memoria della carta.

Gli attributi scelti sono quattro:

- Acquisto minimo: nessun minimo - 50€ - 100€
- Il rivenditore paga una commissione per il servizio: si - no
- Tempo necessario ad eseguire la transazione: immediata (il servizio deve essere sempre on-line) - un giorno (le transazioni sono comunicate tutte a fine giornata).
- Le pratiche possono essere: svolte tutte nel negozio - alcune pratiche compilate in banca.

Attribute	Levels	Utility			Importance		
		Pooled	Group1	Group2	Pooled	Group1	Group2
Placefor paperwork	In-store	0,2321	0,5781	-0.1000	7,61	13,2	5,06
	Bank	-0.2321	-0.5781	0,1			
Commissions	No	0,6454	0,8385	0,46	21,15	19,14	23,27
	Yes	-0.6454	-0.8385	-0.4600			
Time to process	Immediate	0	0	0	37,2	29,85	49,07
	One day	-2.2704	-2.6146	-1.9400			
Minimum purchase	0	0	0	0	34,05	37,81	22,6
	50	-0.9728	-1.8750	-0.1067			
	100	-2.0782	-3.3125	-0.8933			

**Tabella 3.6**

L'analisi è stata eseguita con una full profile su 98 negozi in Europa. Dai dati si sono potuti creare due gruppi: nel primo i rispondenti che preferivano il profilo "sintesi dei livelli con maggiore utilità" e nel secondo gruppo i rimanenti rispondenti.

I risultati sono importanti perché evidenziano due gruppi di rispondenti con preferenze diverse, in cui, oltre a preferire un servizio in "real time" e a nessun costo, si distinguono sul dove svolgere le pratiche e dai differenti valori delle utilità relative.

L'articolo successivo è del 2000 e integra Conjoint Analysis e SERVQUAL per analizzare la percezione della qualità del servizio fornito dalle banche [Oppewal e Vries, 2000]. L'introduzione parla dell'aumentato interesse delle banche nel capire come il cliente percepisca il servizio. Viene spiegato il SERVQUAL, una modalità di indagine tipica dei servizi, in cui alle persone viene fornito un questionario dove si chiede di valutare vari aspetti del servizio: dall'accessibilità, all'estetica, dalla simpatia del personale con cui si è parlato a molte altre voci. L'insieme delle risposte dà un quadro di come il cliente percepisca l'insieme del servizio, l'importanza di questo strumento stà nella capacità di valutare molti aspetti intangibili legati alle sensazioni che il cliente prova, mentre riceve il servizio.

La Conjoint Analysis è introdotta nell'articolo come la modalità statistica che può migliorare i risultati del SERVQUAL trasformando i dati da qualitativi in quantitativi, definendo con precisione le utilità e attribuendo un ordine preciso di preferenza ai vari aspetti del servizio.

Per poter strutturare l'indagine si è ricorsi ad una metodologia chiamata Hierarchical Information Integration (HII), questo perché gli attributi da analizzare erano troppi per un'indagine full profile. L'HII ha permesso di scomporre i vari fattori in quattro insiemi e di mantenerli comunque legati tramite alcuni attributi che sono serviti da ponte tra le parti.

Gli insiemi creati riguardavano l'accessibilità della banca, la competenza del personale, la cura e cortesia e la tangibilità del servizio.

Il test consisteva di 27 profili, per ogni dimensione, ottenuti tramite un piano ortogonale frazionato più altri tre inseriti come holdouts.

Il numero del campione in totale è stato di 200 persone, suddivise tra i quattro test distinti, uno per ogni dimensione. Nella tabella 3.7 sono riportate le dimensioni, i livelli e gli attributi:

- Nella prima dimensione (Accessibilità) i fattori importanti sono stati la funzionalità degli sportelli automatici e la frequenza dell'estratto conto.
- Nella seconda dimensione (Competenza del personale) risultano significativi la capacità di risolvere i problemi e la cura nel trasferire il denaro.
- La terza dimensione (Cura e cortesia) indica l'importanza della cordialità e della cortesia del cassiere allo sportello o al telefono.
- Nell'ultima dimensione (tangibilità) si segnala il tasso di interesse ricevuto, quello pagato e la privacy allo sportello.

I) Accessibilità della banca	II) Competenza del personale	III) Cura e cortesia	IV) Tangibilità
1 Numero di parcheggi -meno dei sufficienti  -sufficienti  -più dei sufficienti	8 Cura nel esecuzione del trasferimento di denaro -quasi mai errori  -a volte un errore	13 Attenzione personale dell'impiegato verso il cliente -alta  -sufficiente  -bassa	22 Estensione dell'assortimento -limitato alla specifica banca(prestiti,mutui..) -esteso con prodotti come assicurazioni...
2 Ore di apertura -solo ore ufficio  -ore ufficio e un sera a settimana -ore ufficio, una sera e il Sabato	9 Iniziativa, es. nel dare consigli. -gli impiegati mostrano molta iniziativa mostrano poca iniziativa	14 Gamma dei servizi computerizzati -piccola  -larga	23 Dimensioni dello spazio del cliente -larga  -sufficiente  -piccola
3 Distanza da percorrere   -meno di 5 minuti  -da 5 a 10 minuti -più di 10 minuti	10 Conoscenza di argomenti particolari o poco richiesti(legislazione, fisco, investimenti) -buona limitata	15 Riconoscimento dal personale  -gli impiegati comunemente mostrano segno di riconoscere i loro clienti - ...a volte -...mai....	24 Arredamento dello spazio del cliente   -minimale  -estesa con articoli come TV, sedie bacheche
4 Operatività dei distributori automatici -mai fuori uso -a volte fuori uso -spesso fuori uso	11 Risoluzione di lamenti -più che adeguata - adeguata -inadeguata	16 Cordialità e cortesia al telefono o allo sportello -impiegati sempre cortesi -solitamente -scortesi e freddi	25 Vestiti del personale -vestiti in uniforme - senza uniforme
5 Coda al distributore   -meno di 2 minuti  -da 2 a 4 minuti -più di 4 minuti	12 Abilità nella comunicazione degli impiegati -buona- giusta-moderata	17 Svolgete gli incarichi  -i compiti sono sempre svolti severamente -non sempre	26 Privacy allo sportello  -buona -sufficiente  -cattiva
6 Coda allo sportello -meno di 2 minuti  -da 2 a 5 minuti -più di 5 minuti		18 Dare avvisi -dando avvisi provano di essere onesti -non sempre dimostrano	27 Interessi pagati sui mutui -1/10 sopra la media  -media del mercato -1/10 sotto la media
7 Frequenza degli estratti-conto -un alla settimana  -uno ogni 2 settimane -uno ogni tre settimane		19 Cura nel gestire i dati dei clienti -la banca è sempre attenta con le informazioni del cliente -.. non essere sempre	
		20 Sicurezza del codice della carta -l'uso non autorizzato della carta è sempre risarcito - non sempre	28 Interessi ricevuti sul capitale -1/10 sopra la media di mercato -media del mercato -1/10 sotto la media
		21 Velocità nel trovare e correggere errori, se sono stati commessi -gli errori sono sempre trovati e corretti -non sempre	

**Tabella 3.7**

L'aggregazione dei dati non modifica i valori delle stime e dimostra come si possano realizzare dei legami tra diversi test ricorrendo ad attributi che esprimano, in sintesi, un giudizio sulle dimensioni non considerate in modo estensivo nel test in esame.

I risultati complessivi sono espressi dalle tabelle sottostanti.

Le conclusioni finali, oltre a dare validità al tipo di metodologia usata, segnalano come per le banche possa essere utile andare a concentrare l'attenzione sulla formazione dei dipendenti e sul motivarli ad essere sempre cortesi e disponibili, e inoltre mantenere un'alta efficienza dei distributori automatici.

Dimension	Importance
III: Accuracy and Friendliness	0.32
II: Competence	0.29
IV: Tangibles	0.23
I: Accessibilty	0.16

**Tabella 3.8**

Classifica dell'importanza degli attributi		
16	Friendly and polite	8,10
11	Settlement of complaints	7,00
4	Money tellers	6,95
28	Interest on savings	5,85
26	Privacy at counter	5,75
27	Interest on loans	5,65
8	Errors in money transfers	5,38
12	Communicative abilities	4,81
7	Frequency of transcripts	4,13
10	Non-specific knowledge	4,13
18	Given advice	3,97
9	Initiatives	3,50
20	Unauthorised card use	3,45
23	Size customer space	3,40
19	Careful with customer data	3,19
2	Opening hours	3,14
13	Personal attention	2,98
22	Assortement	2,93
24	Furnishing of space	2,61
6	Queuing at counter	2,46
5	Queuing at money tellers	2,20
1	Parking place	2,14
17	Meeting commitments	2,04
3	Travel distance	1,88
15	Signs of recognition	1,20
21	Speed of restoring errors	1,10
25	Dress of personnel	0,05
14	Computer services	0,00

**Tabella 3.9**

Nell'articolo più recente del 2003 si analizza come sia possibile mantenere la clientela di studenti nell'ambito del mercato bancario Svedese [Mankila, 2003].

L'analisi parte dalla realtà bancaria che vede gli istituti molto interessati ai giovani, infatti per incentivare la loro adesione ad una banca, molto spesso hanno creato dei pacchetti di servizi a prezzi speciali oppure forniscono dei gadget. Il problema però è come mantenere nel tempo il cliente giovane poiché è sul lungo termine che la banca guadagna, senza considerare che mantenere un cliente costa molto meno che acquisirne uno nuovo. Il mercato è particolarmente difficile perché i giovani cambiano spesso banca, attratti da nuove offerte o migliori servizi. Si può notare come molti approcci tradizionali non funzionino con i giovani ai quali non interessa creare un legame con la banca, né a livello personale né a livello di fedeltà, tramite l'accumulo di punti o premi fedeltà, l'unico punto sensibile per i giovani è la capacità della banca di sviluppare

nuove modalità di servizio tecnologico. L'indagine è stata condotta su 406 studenti di varie facoltà ottenendo 386 questionari validi, il test forniva 16 cartellini ottenuti con un piano ortogonale frazionato più 4 holdouts.

Attributes	Importance	Attributes levels	Part-worth
Current acc	13,51% (13,48)	0,15%+no annual fee	0,3087 (0,1653)
		2,5%+180 SEK fee	-3087 (-0,1653)
Intenationa	18,13% (17,41%)	Yes	1,0813 (1,059)
		No	-1,0813 (-1,059)
Distribution	30.20% (31,67%)	Telephone banking	-0,8439 (-0,9355)
		Internet banking	1,2817 (1,5094)
		Both	0,0405 (-0,1035)
		None	-0,4783 (-0,4704)
Price disco	18,27% (18,63%)	Fees	0,2182 (0,2177)
		Interest rates	-0,6173 (-0,4442)
		Individual	0,3991 (0,2265)
Bank	19,89% (18,82%)	Current bank	0,9477 (1,1022)
		Another major bank	0,1254 (-0,0806)
		Niche bank	-1,0731 (-1,0215)
Model constant			8,2082 (8,1647)

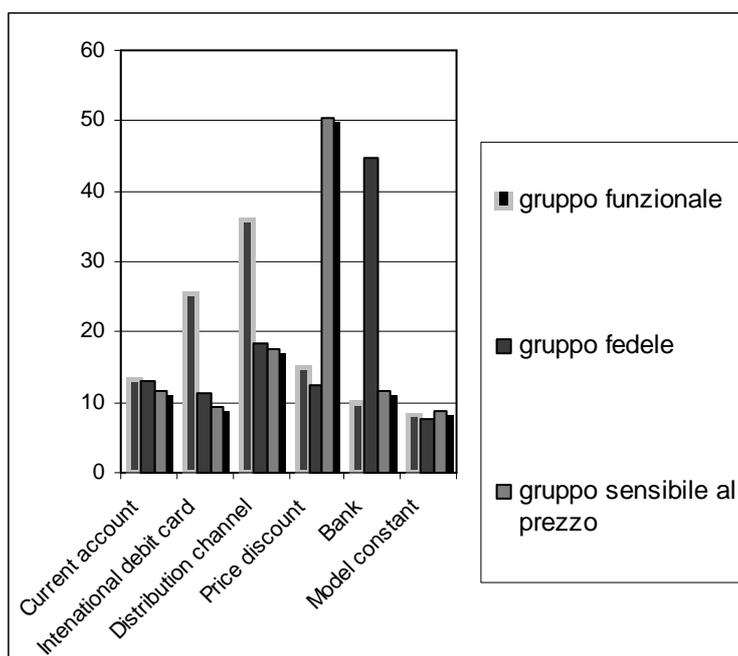
**Tabella 3.10**

I risultati mostrano che il canale distributivo e in particolar modo la possibilità di ottenere servizi tramite internet è il fattore con maggiore importanza. Inoltre la preferenza è quella di restare nella propria banca, di possedere la carta di credito e la possibilità di avere sconti. La variabile che risulta meno importante è quella relativa ai conti correnti e agli assegni circolari.

Dalle analisi si sono potuti costituire tre gruppi di rispondenti:

1. un primo gruppo "funzionale" fortemente sensibile al canale distributivo e interessato ad usare le carte di credito internazionali.
2. Un secondo segmento definito "fedele" interessato alle caratteristiche della banca e in particolare alla propria.
3. Infine un terzo gruppo "sensibile al prezzo" in cui l'importanza è attribuita agli sconti.

Le analisi conclusive indicano che i giovani non considerano molto i pacchetti di servizi creati per loro. Inoltre si riscontra una certa facilità a cambiare banca appena se ne presenta l'opportunità per cercare sempre quella più conveniente. L'aumento della concorrenza spinge le banche a creare un servizio il migliore possibile, che dall'analisi risulta essere, per gli studenti, un pacchetto che includa carte di credito internazionali e l'internet banking. A livello di sconti è molto apprezzata la possibilità di avere agevolazioni proporzionali alla durata del rapporto con la banca.





## Capitolo 4

# ***METODI DI AGGREGAZIONE DEI RISPONDENTI E PROCEDURE NON PARAMETRICHE PER LA VERIFICA D'IPOTESI DI "COERENZA"***

### **4.1 Introduzione**

Questo capitolo è dedicato alla descrizione del metodo di combinazione non parametrica "NPC ranking" [Pesarin e Lago, 1998]; esso rappresenta una buona soluzione a problemi di classificazione di variabili, in base a giudizi raccolti tramite interviste su di un campione (unità statistiche). Inoltre, per confrontare i risultati ottenuti dall'utilizzo di questo metodo rispetto alla media aritmetica, attualmente usata in letteratura, si utilizzeranno l'indice di Spearman  $I_S$  e l'indicatore di permutazione sulle graduatorie  $I_{PG}$ .

### **4.2 Presentazione del problema**

Quando si confrontano tra loro soggetti di diversa natura (qualità di un prodotto, comportamenti, persone, aziende, prodotti, servizi, ecc...), spesso si desidera ottenere una classifica degli attributi scelti secondo l'interesse dell'indagine. Se l'analisi è costruita sulla base di una sola caratteristica (ad.es. il prezzo), il risultato è scontato ed immediato non essendoci problemi connessi all'ordinamento. Il problema nasce quando ci si trova a lavorare con più di una caratteristica simultaneamente, come ad esempio il prezzo, la gradazione di colore e la dimensione. In questo caso si costruiscono tante classifiche quante sono le unità statistiche che formano il campione, ottenendo così un

insieme di graduatorie parziali che saranno combinate in modo da ordinare tutti i punti in una classifica globale che dovrà tener conto di tutte le caratteristiche considerate.

Il metodo statistico che verrà qui esaminato si prefigge questo obiettivo. Trattandosi di una procedura non parametrica non richiede né la conoscenza della distribuzione statistica delle variabili né la loro dipendenza strutturale. Inoltre, un secondo vantaggio di questa metodologia è rappresentato dal richiedere assunzioni deboli e facilmente controllabili sui dati. Per questi motivi la procedura non parametrica che si andrà a descrivere gode di un'ampia potenzialità applicativa.

### 4.3 Condizioni di applicabilità del metodo

Si consideri il caso in cui si abbiano  $k$  cartellini (insieme di caratteristiche) e  $n$  rispondenti (unità statistiche). Il problema che si vuole risolvere è di costruire una graduatoria finale dei  $k$  cartellini combinando le  $n$  graduatorie parziali relative a ciascun rispondente. Per fare questo si è scelto di utilizzare sia l'NPC ranking che la media per poi porli a confronto.

L'esempio illustrato può essere descritto nel seguente modo:

- numero di rispondenti  $n = 5$ ,
- numero di attributi  $p = 2$ ,
- numero di livelli (+ e -)  $l = 2$ ,
- di conseguenza il numero dei cartellini sarà  $k = (l^p = 2^2) = 4$  come illustrato nella tabella 4.1.

Ai rispondenti sarà richiesto di assegnare ad ogni cartellino un valore da 0 a 10 (dove 10 è meglio), mentre le due graduatorie finali NPC e Media sono espresse in ranghi (dove 4 è meglio). Nel seguito (al paragrafo 4.4) verrà spiegato come ottenere i ranghi.

	<b>A</b>	<b>B</b>	<b>Risp_1</b>	<b>Risp_2</b>	<i>Risp_3</i>	<b>Risp_4</b>	<b>Risp_5</b>	<b>NPC</b>	<b>Media</b>
<b>Cart1</b>	+	+	8	5	7	4	9	<b>4</b>	<b>4</b>
<b>Cart2</b>	+	-	6	7	2	8	7	<b>3</b>	<b>3</b>
<b>Cart3</b>	-	+	5	3	6	2	5	<b>1</b>	<b>2</b>
<b>Cart4</b>	-	-	2	4	4	9	1	<b>2</b>	<b>1</b>

**Tabella 4.1**

Risp<sub>j</sub> con  $j = 1, \dots, 5$  è il j-esimo intervistato, Cart<sub>i</sub> con  $i=1, \dots, 4$  è l'i-esimo cartellino

Le condizioni per poter applicare questo metodo sono sostanzialmente due:

- a) l'ordinamento di tutti i cartellini deve rispondere ad un solo senso (crescente o decrescente) in modo che, ad esempio, a valori elevati di ogni cartellino corrisponda una posizione di alta classifica (grande è meglio) per ciascuna unità statistica.
- b) i **k** cartellini non devono essere degeneri cioè non devono assumere un solo valore con probabilità 1. Infatti, se così fosse, si otterrebbe una graduatoria con tutti ex-equò, che non produrrebbe nessuna informazione utile.

Come si può vedere le condizioni poste sono poco restrittive e facilmente controllabili, questo consente di applicare la metodologia a numerose situazioni reali.

#### **4.4 Metodologia della procedura**

In una prima fase verranno costruite le classifiche parziali che, combinate tra loro tramite la funzione di combinazione di Fisher, formeranno la graduatoria globale.

#### 4.4.1 Costruzione delle classifiche parziali

Al fine di costruire le classifiche parziali risulta innanzitutto necessario definire un opportuno criterio di valutazione dei cartellini, nel senso che a valori grandi del cartellino  $i$ -esimo per l'unità statistica  $h$ -esima corrisponda una posizione di alta classifica. Definiamo  $x_{hi}$  il valore assunto dal cartellino  $i$ -esimo per l' $h$ -esima unità statistica in una matrice  $X$  di dimensioni  $k \times n$ , con  $i = 1, \dots, k$  e  $h = 1, \dots, n$ .

Calcoliamo per ogni  $x_{hi}$  il seguente rapporto:

$$R_{hi} = \frac{\#(X_{hi} \geq X_{hj}) + 0,5}{k + 1}$$

con  $i, j = 1, \dots, k$ ;  $h = 1, \dots, n$  e  $\#$  è il numero di volte in cui si verifica la condizione espressa tra parentesi, cioè il rango.

Così facendo si ottiene una matrice  $R$  di dimensione  $k \times n$ , i cui elementi sono gli  $R_{hi}$  con  $i = 1, \dots, k$  e  $h = 1, \dots, n$ . Le colonne di  $R$ , ordinate in modo decrescente al loro interno, formano le classifiche parziali.

#### 4.4.2 Costruzione della classifica finale

Per arrivare ad ottenere una sola classifica globale è necessario applicare agli elementi di ciascuna riga di questa matrice, prima di essere ordinata, la funzione di combinazione di Fisher:

$$Z_i = - \sum_{h=1}^n W_h * \log(1 - R_{hi})$$

dove  $w_h$  è il peso dato all'h-esimo rispondente con  $h=1, \dots, n$  e con  $\sum_{h=1}^n w_h = 1$ , mentre gli  $R_{hi}$  corrispondono agli elementi della matrice  $R$ .

In questo modo tutte le  $k$  informazioni di ogni singola riga vengono riassunte in un unico valore  $z_i$ . La funzione di combinazione di Fisher è reale continua, non decrescente e non degenere, che associa ad un vettore in  $\mathbb{R}^n$  un numero di  $\mathbb{R}^1$ . Inoltre, essa soddisfa ai seguenti requisiti: l'invarianza rispetto a permutazioni delle variabili, nel senso che se anche si cambia l'ordine d'imputazione delle variabili il risultato non varia; ciascun argomento deve essere non decrescente ed infine la funzione raggiunge il suo valore massimo quando almeno uno degli argomenti assume valore unitario. Nell'espressione sopra riportata compare  $w_h$ , questo rappresenta il peso opportunamente scelto in funzione dell'importanza del singolo rispondente. Ad esempio se un'azienda è interessata a lanciare nel mercato un prodotto più adatto alle esigenze di un manager, la risposta di quest'ultimo avrà un peso superiore (es.:  $w_{\text{manager}}=0,9$ ) rispetto a quella di un semplice impiegato (es.:  $w_{\text{impiegato}}=0,1$ ).

L'ultimo passaggio per ottenere la classifica finale consiste nel confrontare i diversi  $z_i$  tra loro nel seguente modo:

$$G_f(i) = \frac{\#(Z_i \geq Z_j)}{k}$$

Si ottiene in questo modo un vettore colonna di valori che, se ordinati in modo decrescente, forniscono la classifica finale.

## 4.5 Confronto tra graduatorie

Per confrontare i risultati ottenuti dall'utilizzo del NPC ranking rispetto alla media aritmetica, si utilizzano l'indice di Spearman  $I_S$  e l'indicatore  $I_{PG}$ .

#### 4.5.1 L'indice di Spearman $I_S$ ed il test TS

Per verificare se il metodo simula in modo accettabile il processo decisionale preesistente ed, inoltre, al fine di calcolare la correlazione tra graduatorie, risulta necessario il ricorso ad un indice capace di sintetizzare numericamente il grado di somiglianza tra due classifiche.

Una possibile soluzione formulazione è data dalla seguente formulazione dell'indice di Spearman  $I_S$  o indice di cograduazione (correlazione tra ranghi):

$$I_S = \frac{3 * \sum_{i=1}^k (R_i - R'_i)^2}{k * (k^2 - 1)}$$

dove  $R_i$  è il rango dell' $i$ -esimo valore della classifica A, mentre  $R'_i$  è il rango dell' $i$ -esimo valore della classifica B. L'indice  $I_S$  assume valori compresi nell'intervallo  $[0,1]$ , più si avvicina a 0 più la graduatoria osservata è simile alla classifica di riferimento.

Successivamente, dopo aver calcolato gli indici di Spearman  $S^*$  su tutte le possibili permutazioni della graduatoria di riferimento, viene calcolato il test di Spearman (TS) nel seguente modo:

$$TS = \frac{\#(S^{oss} \geq S^*)}{B}$$

dove  $S^{oss}$  rappresenta l'indice di Spearman calcolato sulla graduatoria osservata,  $S^*$  sono gli indici di Spearman calcolati su tutte le permutazioni e B è il numero di tutte le possibili permutazioni del vettore di riferimento.

#### 4.5.2 L'indicatore di permutazione sulle graduatorie $I_{PG}$ ed il test TPG

Il Test di permutazione misura la coerenza tra un'ipotetica graduatoria di riferimento e la graduatoria ottenuta con l'NPC ranking o con la media. In altre parole, si desidera valutare quanto le due classifiche portino a risultati simili o, meglio, non si contraddicano tra loro (se  $x_1 < x_2$  secondo la graduatoria A allora il confronto  $x_1 < x_2$  deve valere anche per la graduatoria B).

Formalmente questo può essere riassunto tramite il seguente sistema d'ipotesi:

$$\begin{cases} H_0: \text{assenza di ordinamento} \\ H_1: \text{presenza dell'ordinamento } \{rk(x_1)^{rif}, \dots, rk(x_n)^{rif}\}. \end{cases}$$

e la statistica test di riferimento è la seguente:

$$T(\{rk(x_1)^{oss}, \dots, rk(x_n)^{oss}\}) = \sum_{1 \leq i < j \leq n} [ k(ij) + k(ij) * L(ij) + k(ij) * H(ij) ]$$

dove

$K(ij)$  assume valore: 1 quando  $(rk(x_i)^{rif} > rk(x_j)^{rif})$  e  $(rk(x_i)^{oss} < rk(x_j)^{oss})$   
o  $(rk(x_i)^{rif} < rk(x_j)^{rif})$  e  $(rk(x_i)^{oss} > rk(x_j)^{oss})$ ,

0 altrimenti.

$$L(ij) = |rk(x_i)^{rif} - rk(x_j)^{rif}| - 1$$

$$H(ij) = |rk(x_i)^{oss} - rk(x_j)^{oss}| - 1$$

Si osserva come  $K(ij)$  valuti la coerenza tra le graduatorie,  $L(ij)$  misuri la distanza tra le osservazioni nella graduatoria di riferimento e  $H(ij)$  la distanza tra le osservazioni nella classifica osservata. Più la statistica  $T$  si avvicina allo 0 più la graduatoria ottenuta dal metodo implementato usato risulterà coerente con il principio attraverso il quale si è

ottenuta la classifica di riferimento. Ora, si suppone, a titolo di esempio, di considerare 2 attributi A e B su due livelli (+ e -) come riportato in tabella 4.2:

	<b>A</b>	<b>B</b>	<b>risposte</b>	<b>ordinamento di rif.</b>
<b>Cart.1</b>	+	+	$y_1 \sim N(\mu_1, \sigma^2)$	4
<b>Cart.2</b>	+	-	$y_2 \sim N(\mu_2, \sigma^2)$	3
<b>Cart.3</b>	-	+	$y_3 \sim N(\mu_3, \sigma^2)$	2
<b>Cart.4</b>	-	-	$y_4 \sim N(\mu_4, \sigma^2)$	1

**Tabella 4.2**

ed ipotizzando questa condizione:  $\mu_1 > \mu_2 > \mu_3 > \mu_4$  si otterrà un ordinamento per ranghi (rk) del tipo: 4 3 2 1.

In questo caso la verifica di ipotesi rispetto all'ordinamento di riferimento è la seguente:

$$\begin{cases} H_0: \text{assenza dell'ordinamento relativo alla graduatoria di rif.} \\ H_1: \text{presenza dell'ordinamento di rif.} \end{cases}$$

Ipotizzando di aver ottenuto una classifica finale di questo genere (tabella 4.3):

	<b>class.finale oss.</b>	<b>class.di rif.</b>
<b>Cart.1</b>	2	4
<b>Cart.2</b>	3	3
<b>Cart.3</b>	1	2
<b>Cart.4</b>	4	1

**Tabella 4.3**

tutti i possibili confronti a coppie sono:  $\binom{4}{2} = 6$  e le singole coppie risultano:

(cart1, cart2) (cart1, cart3) (cart1, cart4) (cart2, cart3) (cart2, cart4) (cart3, cart4).

La statistica test si calcola nel seguente modo :

$$\mathbf{T}(\{\text{rk}(x_1)^{\text{oss}}, \dots, \text{rk}(x_n)^{\text{oss}}\}) = \sum_{1 \leq i \leq j \leq n} [k(ij) + k(ij) * L(ij) + k(ij) * H(ij)]$$

dove:  
**K (ij)** assume valore: 1 quando  $(rk(x_i)^{rif} > rk(x_j)^{rif})$  e  $(rk(x_i)^{oss} < rk(x_j)^{oss})$   
o  $(rk(x_i)^{rif} < rk(x_j)^{rif})$  e  $(rk(x_i)^{oss} > rk(x_j)^{oss})$ ,  
0 altrimenti.

**K**

1	0	1
0	0	1
0	0	1

**L (ij)** =  $|rk(x_i)^{rif} - rk(x_j)^{rif}| - 1$

**L**

0	1	2
0	0	1
0	0	0

**H (ij)** =  $|rk(x_i)^{oss} - rk(x_j)^{oss}| - 1$

**H**

0	0	1
0	1	0
0	0	2

$$\begin{aligned} \text{Quindi } T &= \sum_{1 \leq i \leq j \leq n} [k(ij) + \sum k(ij) * L(ij) + \sum k(ij) * H(ij)] = \\ &= 4 + 3 + 3 = 10 \end{aligned}$$

Per normalizzare la statistica T si calcola il suo massimo prendendo in considerazione la graduatoria osservata più distante (1 2 3 4) rispetto a quella di riferimento (4 3 2 1) nel modo seguente:

$$\mathbf{T}_{\max} = \frac{k * (k^2 - 1)}{3} - \frac{k * (k - 1)}{2}$$

che nell'esempio diventa:  $\frac{4 * (16 - 1)}{3} - \frac{4 * (4 - 1)}{2} = 14$

Si dimostra nel seguente modo come si è ottenuto  $\mathbf{T}_{\max}$ :

Sapendo che:  $\sum_{i=1}^k i = \frac{k}{2} * (k + 1)$  e  $\sum_{i=1}^k i^2 = \frac{k}{6} * (k + 1) * (2k + 1)$

ed essendo:

$$\mathbf{L}(ij) = | \text{rk}(x_i)^{\text{rif}} - \text{rk}(x_j)^{\text{rif}} | - 1$$

$$\Rightarrow \sum_{i < j} | \text{rk}(x_i)^{\text{rif}} - \text{rk}(x_j)^{\text{rif}} | = \begin{array}{|c|c|c|} \hline \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \hline \mathbf{0} & \mathbf{1} & \mathbf{2} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \hline \end{array} =$$

$$= 3*1 + 2*2 + 1*3 = \sum_{i=1}^{k-1} i * (k - i) = k * \sum_{i=1}^{k-1} i - \sum_{i=1}^{k-1} i^2 = \frac{k^2}{2} * (k + 1) - \frac{k}{6} * (k + 1) * (2k + 1) =$$

$$= (k+1) * \frac{3k^2 - k * (2k + 1)}{6} = \frac{(k + 1) * (k^2 - k)}{6} = k * (k+1) * (k-1) =$$

$$= \frac{k}{6} * (k^2 - 1).$$

Sapendo che nel caso peggiore  $\mathbf{L}(ij)$  è uguale a  $\mathbf{H}(ij)$  il massimo valore che può assumere il Test è:

$$\mathbf{T}_{\max} = \sum_{1 \leq i < j \leq n} [ k(ij) + \sum k(ij) * \mathbf{L}(ij) + \sum k(ij) * \mathbf{H}(ij) ] =$$

$$= \binom{k}{2} + \frac{k}{6} * (k^2 - 1) - \binom{k}{2} + \frac{k}{6} * (k^2 - 1) - \binom{k}{2} =$$

$$= \frac{k * (k^2 - 1)}{3} - \frac{k * (k - 1)}{2} .$$

Quindi, normalizzando la statistica test precedente, si ottiene l'indicatore  $I_{PG}$ :

$$I_{PG} = \frac{T}{\frac{k * (k^2 - 1)}{3} - \frac{k * (k - 1)}{2}}$$

che assume valori compresi nell'intervallo [0,1]: più l'indice si avvicina a 0 più la graduatoria osservata è simile alla classifica di riferimento.

Successivamente, dopo aver calcolato l' $I_{PG} T^*$  su tutte le possibili permutazioni della graduatoria di riferimento, si calcola il test di permutazione sulle graduatorie (TPG) nel seguente modo:

$$TPG = \frac{\#(T^{oss} \geq T^*)}{B}$$

dove  $T^{oss}$  è l' $I_{PG}$  calcolato sulla graduatoria osservata,  $T^*$  sono gli  $I_{PG}$  calcolati su tutte le permutazioni e B è il numero di tutte le possibili permutazioni del vettore di riferimento.

## 4.6 I passi dell'algoritmo

### Passo 1.

Come prima operazione si raggruppano le risposte degli intervistati in una matrice X di dimensione  $k \times n$  ( $k = 5$  cartellini,  $n = 6$  rispondenti) come nell'esempio riportato in tabella 4.4.

var \ Resp	Risp_1	Risp_2	Risp_3	Risp_4	Risp_5	Risp_6
<i>Cart.1</i>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>14</sub>	X <sub>15</sub>	X <sub>16</sub>
<b>Cart.2</b>	X <sub>21</sub>	X <sub>22</sub>	X <sub>23</sub>	X <sub>24</sub>	X <sub>25</sub>	X <sub>26</sub>
<b>Cart.3</b>	X <sub>31</sub>	X <sub>32</sub>	X <sub>33</sub>	X <sub>34</sub>	X <sub>35</sub>	X <sub>36</sub>
<b>Cart.4</b>	X <sub>41</sub>	X <sub>42</sub>	X <sub>43</sub>	X <sub>44</sub>	X <sub>45</sub>	X <sub>46</sub>
<b>Cart.5</b>	X <sub>51</sub>	X <sub>52</sub>	X <sub>53</sub>	X <sub>54</sub>	X <sub>55</sub>	X <sub>56</sub>

**Tabella 4.4**

Risp<sub>j</sub> con j = 1,...,6 è il j-esimo intervistato, Cart.i con i = 1,...,5 è l'i-esimo cartellino e x<sub>ij</sub> è il valore dato dal j-esimo rispondente all'i-esimo cartellino.

**Passo 2.**

Si calcola il rango di ogni colonna della matrice X, cioè il # (x<sub>i</sub> ≥ x<sub>j</sub>) ottenendo così una nuova matrice (**kxn**) in cui gli elementi iniziali sono sostituiti dai rispettivi ranghi.

**Passo 3.**

Si calcola il seguente rapporto:

$$R_{hi} = \frac{\#(X_{hi} \geq X_{hj}) + 0,5}{k + 1}$$

Così facendo si ottiene una matrice R di dimensione kxn, i cui elementi sono gli **R<sub>hi</sub>** con i=1,...,k e h=1,...,n. Le colonne di R, ordinate in modo decrescente al loro interno, formano le classifiche parziali.

**Passo 4.**

Si applica ad ogni riga della matrice R la funzione di combinazione di Fisher:

$$Z_i = - \sum_{h=1}^n W_h * \log(1 - R_{hi})$$

ottenendo il vettore degli Z<sub>i</sub>.

**Passo 5.**

Come ultimo passaggio dell'algoritmo si confrontano gli  $Z_i$  con il metodo dei ranghi:

$$G_j(i) = \frac{\#(Z_i \geq Z_j)}{k}$$

Così facendo si ottiene la classifica finale.

**Passo 6.**

Il passo successivo è quello di calcolare la classifica finale attraverso la media aritmetica, per poi porla a confronto con l'NPC e valutare quale tra i due risultati il migliore. Quindi si, calcola la media di ogni riga della matrice iniziale X e si applica il metodo dei ranghi; così facendo si ottiene la classifica finale.

**Passo 7.**

Per effettuare il confronto occorre applicare l'indice di Spearman  $I_S$  alle due classifiche ottenute precedentemente:

$$I_S = \frac{3 * \sum_{i=1}^k (R_i - R'_i)^2}{k * (k^2 - 1)}$$

Più l'indice si avvicina ad uno più la classifica finale si allontana da quella reale.

**Passo 8.**

Si calcola, per entrambe le classifiche, la seguente statistica test :

$$T(\{rk(x_1)^{OSS}, \dots, rk(x_n)^{OSS}\}) = \sum_{1 \leq i \leq j \leq n} [ K(ij) + k(ij) * L(ij) + k(ij) * H(ij) ]$$

dove

$K(ij)$  assume valore 1 quando  $(rk(x_i)^{rif} > rk(x_j)^{rif})$  e  $(rk(x_i)^{OSS} < rk(x_j)^{OSS})$   
o  $(rk(x_i)^{rif} < rk(x_j)^{rif})$  e  $(rk(x_i)^{OSS} > rk(x_j)^{OSS})$ ,

0 altrimenti.

$$L(ij) = | \text{rk}(x_i)^{\text{rif}} - \text{rk}(x_j)^{\text{rif}} | - 1$$

$$H(ij) = | \text{rk}(x_i)^{\text{oss}} - \text{rk}(x_j)^{\text{oss}} | - 1$$

### **Passo 9.**

Infine viene calcolato l'indicatore  $I_{PG}$ :

$$I_{PG} = \frac{T}{\frac{k * (k^2 - 1)}{3} - \frac{k * (k - 1)}{2}} ;$$

Più l'indice risulta prossimo ad uno più la classifica finale si allontana da quella reale.

## **4.7 Implementazione in SAS**

### **4.7.1 Introduzione**

Per creare la macro si necessita di un vettore colonna di pesi (con un numero di righe uguale al numero di rispondenti) e di un vettore colonna con la media di riferimento (mediag):

-come prima operazione si importa il data set dei pesi-

```
libname pesi 'C:\percorso';
```

```
data pesi;
```

```
infile 'C:\percorso\pesi1.txt';
```

```
input w1;
```

```
run;
```

-poi si importa il data set della media di riferimento-

```
libname mediag 'C:\percorso';
```

```

data mediag;
infile 'C:\percorso\medie.txt';
input m1;
run;

```

Si entra nel linguaggio macro chiamando la procedura **graduatoria** assegnando a **w** il vettore pesi, a **u** il vettore mediag, a **n** il numero di rispondenti ed a **k** il numero di cartellini:

```
%macro graduatoria(w,u,n,k);
```

Ai fini della simulazione si generano dei dati casuali che corrispondono alle varie risposte degli intervistati, mentre, nel caso in cui si sia in possesso di un data set con le risposte, si utilizzerà quest'ultimo senza generare alcun dato:

-si utilizza l'ambiente iml che permette di creare matrici e di effettuare operazioni tra loro-

```
proc iml;
```

- si richiama il data set mediag chiamandolo &u perché siamo all'interno di una procedura macro-

```
use &u;
```

- si definiscono x gli elementi del nostro data set-

```
read all var _num_ into x;
```

-si crea una matrice base kxn di tutti 1 nella quale verranno inseriti i dati generati-

```
ncas=j(&k,&n,1);
```

-si inizializzano due cicli for -

```
do i=1 to &k;
```

```
do j=1 to &n;
```

-si sostituiscono gli 1 della matrice precedentemente creata con i numeri generati casualmente dalla distribuzione normale –

```
ncas[i,j]=sqrt(1)*normal(seed)+x[i];
```

-si chiude il primo ciclo-

```

end;
-si chiude il secondo ciclo-
end;
-si crea un data set di nome casuale che avrà al suo interno la matrice di numeri casuali-
create ncasuale from ncas;
append from ncas;
-si esce dall'ambiente iml-
quit;
-si stampa il dataset creato-
proc print data=ncasuale;
run;

```

#### 4.7.2 Graduatoria ottenuta con l'NPC ranking

Si calcola il rango di ogni colonna della matrice  $X$ , cioè' il  $\#(x_i \geq x_j)$  ottenendo così una nuova matrice ( $k \times n$ ) in cui gli elementi iniziali sono sostituiti dai rispettivi ranghi.

```

proc rank data=ncasuale ties=high out=rank;
run;

```

Si calcola il seguente rapporto:

$$R_{hi} = \frac{\#(X_{hi} \geq X_{hj}) + 0,5}{k + 1}$$

-si definisce un nuovo data set -

```
data prv;
```

-si utilizza il data set rank creato precedentemente con la procedura proc rank-  
set rank;

-si definiscono tanti vettori quante sono le colonne della matrice del data set rank per poi sostituire al loro interno il risultato dell'operazione sotto riportata-

```
array prov(&n) col1-col&n;
```

array new(&n) a1-a&n;

-si inizializza un ciclo for-

```
do i=1 to &n;
```

-si esegue su ogni elemento della matrice X la seguente operazione-

```
new(i)=(prov(i)+0.5)/(&k+1);
```

-si chiude il ciclo-

```
end;
```

-si definisce un nuovo data set-

```
data R;
```

-si utilizza il data set prv prendendo solo le colonne a1-a&n-

```
set prv (keep=a1-a&n);
```

```
run;
```

Si applica ad ogni riga della matrice R la funzione di combinazione di Fisher:

$$Z_i = - \sum_{h=1}^n W_h * \log(1 - R_{hi})$$

```
data fisher;
```

-si utilizza il data set R-

```
set R;
```

-si definiscono tanti vettori quante sono le colonne di R per poi sostituirci al loro interno il risultato dell'operazione sotto riportata-

```
array prov(&n) a1-a&n;
```

```
array new(&n) b1-b&n;
```

-si inizializza un ciclo for-

```
do i=1 to &n;
```

-si esegue su ogni elemento della matrice R la seguente operazione-

```
new(i) = -log(1-prov(i));
```

-si chiude il ciclo for-

*end;*

*run;*

-si definisce un nuovo data set-

*data Z;*

-si prendono solo le colonne b1-b&n del data set fisher-

*set fisher (keep=b1-b&n);*

*run;*

Si calcola il vettore degli  $Z_i$ :

-si utilizza l'ambiente iml-

*proc iml;*

-si utilizza il data set Z-

*use Z;*

-si definiscono x gli elementi del nostro data set-

*read all var \_num\_ into x;*

-si utilizza il vettore dei pesi-

*use &w;*

-si definiscono y gli elementi di w-

*read all var \_num\_ into y;*

-si moltiplica  $x * y$  così da ottenere il vettore degli  $Z_i$ -

$h = x * y;$

-si definisce re il data set così ottenuto-

*create re from h;*

*append from h;*

*quit;*

*run;*

Come ultimo passaggio dell'algoritmo si confrontano gli  $Z_i$  attraverso il metodo dei ranghi:

$$G_f(i) = \frac{\#(Z_i \geq Z_j)}{k}$$

```
proc rank data=re ties=high out=rank1;
run;
```

Dividendo il vettore degli  $Z_i$ , ordinato secondo la procedura rank, per  $k$  si ottiene la graduatoria finale normalizzata:

```
data clasfin;
set rank1;
-si divide il vettore degli  $Z_i$  per  $k$ -
graduatoria=(COL1)/(&k);
-si crea il data set finale cha ha un'unica colonna chiamata graduatoria-
data finale;
set clasfin(keep = graduatoria);
run;
```

### 4.7.3 Graduatoria ottenuta con la media

Si calcola la media di ogni riga della matrice iniziale  $X$  e si applica il metodo dei ranghi. Così facendo si ottiene la classifica finale.

```
proc iml;
-si utilizza il data set ncasuale dei dati creati-
use ncasuale;
-si definisce x la matrice kxn del data set ncasuale-
read all var _num_ into x;
-si utilizza il data set w (nx1) dei pesi-
use &w;
read all var _num_ into y;
-si calcola la somma di ogni riga della matrice del data set ncasuale-
m=(x*y);
-si crea il data set med formato da una colonna Col1-
create med from m;
```

```

append from m;
quit;
run;
data media;
set med;
-si calcola la media aritmetica di ogni elemento della colonna Col1-
classifica=(COL1)/(&n);
-si crea il data set medie prendendo solo la colonna classifica-
data medie;
set media(keep = classifica);
-si applicano i ranghi al vettore delle medie -
proc rank data=medie ties=high out=rankm;
run;

```

#### 4.7.4 Indice $I_S$ per l'NPC ranking

Per calcolare l'indice di Spearman  $I_S$ :

$$I_S = \frac{3 * \sum_{i=1}^k (R_i - R'_i)^2}{k * (k^2 - 1)}$$

occorre confrontare il vettore della media di riferimento(&u) con il vettore ottenuto con l'NPC ranking di graduatorie ( rank1 ) :

```

proc iml;
-si richiama il vettore della media di riferimento e si definisce x -
use &u;
read all var _num_ into x;

```

-si richiama il vettore della classifica finale ottenuto con il metodo delle combinazioni non parametriche e lo si definisce y-

```
use rank1;
```

```
read all var _num_ into y;
```

-si calcola la differenza per riga degli elementi dei 2 vettori-

```
s=(y-x);
```

-si crea il data set s2 formato dal vettore delle differenze-

```
create s2 from s;
```

```
append from s;
```

```
quit;
```

```
run;
```

```
proc iml;
```

- si richiama il data set s2 e lo si definisce x-

```
use s2;
```

```
read all var _num_ into x;
```

-si definisce il vettore x come dif2-

```
dif2=x;
```

-si crea una matrice kx1 di tutti 0-

```
quadr2=j(&k,1,0);
```

-si inizializza un ciclo for-

```
do i=1 to &k;
```

-si sostituisce agli 0 della matrice quadr2 il quadrato delle differenze-

```
quadr2[i,1]=dif2[i]*dif2[i];
```

```
end;
```

-si crea il data set dquad2 formato dal vettore con le differenze al quadrato-

```
create dquad2 from quadr2;
```

```
append from quadr2;
```

```
quit;
```

```
run;
```

```
proc iml;
```

-si utilizza il data set dquad2 e lo si definisce q2 -

```
use dquad2;
```

```

read all var _num_ into q2;
-si calcola il numeratore della formula di Spearman-
somap=3*sum(q2);
-si crea il data set s2a contenente il numeratore di Spearman-
create s2a from sommap;
append from sommap;
quit;
run;
-si definisce un nuovo data set sp2-
data sp2;
set s2a;
-calcolo finale dell'indice di Spearman-
Spp=(COL1/(&k*(&k*&k-1)));
data spearmanp;
set sp2(keep =Spp);
-si stampa il valore ottenuto dall'indice di Spearman-
proc print data=spearmanp;
run;

```

#### 4.7.5 Indice $I_S$ per la media

L'indice di Spearman  $I_S$  viene calcolato confrontando il vettore della media di riferimento (&u) con il vettore ottenuto con la media aritmetica (rankm):

$$I_S = \frac{3 * \sum_{i=1}^k (R_i - R'_i)^2}{k * (k^2 - 1)}$$

```

proc iml;
-si richiama il vettore della media di riferimento chiamandolo x-
use &u;

```

```
read all var _num_ into x;
```

-si richiama il vettore della classifica finale ottenuto con il metodo della media chiamandolo y-

```
use rankm;
```

```
read all var _num_ into y;
```

-si calcola la differenza per riga degli elementi dei 2 vettori-

```
s=(y-x);
```

-si crea il data set s1 formato dal vettore delle differenze-

```
create s1 from s;
```

```
append from s;
```

```
quit;
```

```
run;
```

```
proc iml;
```

```
use s1;
```

```
read all var _num_ into x;
```

-si definisce il vettore x come dif1-

```
dif=x;
```

-si crea una matrice kx1 di tutti 0-

```
quadr=j(&k,1,0);
```

-si inizializza il ciclo for-

```
do i=1 to &k;
```

-si sostituisce agli 0 della matrice il quadrato delle differenze-

```
quadr[i,1]=dif[i]*dif[i];
```

-chiudiamo il ciclo-

```
end;
```

-si crea il data set dquad formato dal vettore con le differenze al quadrato-

```
create dquad from quadr;
```

```
append from quadr;
```

```
quit;
```

```
run;
```

```
proc iml;
```

-si utilizza il data set dquad e lo si definisce q1-

```

use dquad;
read all var _num_ into q1;
-si calcola il numeratore della formula di Spearman-
sommam=3*sum(q1);
-si crea il data set s1a contenente il numeratore di Speraman-
create s1a from sommam;
append from sommam;
quit;
run;
data sp1;
set s1a;
-calcolo finale dell'indice di Spearman-
Spm=(COL1/(&k*(&k*&k-1)));
data spearmanm;
set sp1(keep =Spm);
-si stampa il valore ottenuto dall'indice di Spearman-
proc print data=spearmanm;
run;

```

#### 4.7.6 Indicatore $I_{PG}$ per l'NPC ranking

Si applica la statistica test al vettore della media di riferimento ( $\mu$ ) ed al vettore ottenuto con l'NPC ranking (rank1):

$$\mathbf{T}(\{\text{rk}(x_1)^{\text{OSS}}, \dots, \text{rk}(x_n)^{\text{OSS}}\}) = \sum_{1 \leq i \leq j \leq n} [ K(ij) + k(ij) * L(ij) + k(ij) * H(ij) ]$$

-si calcola la matrice K-

$K(ij)$  assume valore 1 quando  $(rk(x_i)^{rif} > rk(x_j)^{rif})$  e  $(rk(x_i)^{oss} < rk(x_j)^{oss})$   
o  $(rk(x_i)^{rif} < rk(x_j)^{rif})$  e  $(rk(x_i)^{oss} > rk(x_j)^{oss})$ ,  
0 altrimenti.

*proc iml;*

-si utilizza il vettore ottenuto dal metodo delle combinazioni non parametriche di graduatorie (rank1) chiamandolo x-

*use rank1;*

*read all var \_num\_ into x;*

-si crea una matrice k-1 x k-1 di tutti 0-

*matp=j(&k-1,&k-1,0);*

-si inizializzano due cicli for-

*do i=1 to &k-1;*

*do j=i+1 to &k;*

-se l'elemento i del vettore x è > dell'elemento j del vettore x si inserisce 1 nella matrice matp nel posto [i,j-1]-

*if x[i]>x[j]*

*then matp[i,j-1]=1;*

-chiusura primo ciclo-

*end;*

-chiusura secondo ciclo-

*end;*

-si crea il data set Kp contenente la matrice K di 0 e 1-

*create Kp from matp;*

*append from matp;*

*quit;*

*run;*

-si crea la matrice L-

$$L(ij) = | rk(x_i)^{rif} - rk(x_j)^{rif} | - 1$$

*proc iml;*

-si richiama il vettore della media di riferimento chiamandolo x-

*use &u;*

*read all var \_num\_ into x;*

-si crea una matrice k-1xk-1 di tutti 0-

*matlp=j(&k-1,&k-1,0);*

-si inizializzano due cicli for-

*do i=1 to &k-1;*

*do j=i+1 to &k;*

-si calcola la distanza in valore assoluto -1 tra gli elementi del vettore x-

*matlp[i,j-1]=abs(x[i]-x[j])-1;*

-chiusura primo ciclo-

*end;*

-chiusura secondo ciclo-

*end;*

-si crea il data set Lp contenente la matrice L -

*create Lp from matlp;*

*append from matlp;*

*quit;*

*run;*

-si crea la matrice H-

$$H(ij) = | rk(x_i)^{oss} - rk(x_j)^{oss} | - 1$$

*proc iml;*

-si utilizza il data set rank1 chiamandolo x-

*use rank1 ;*

*read all var \_num\_ into x;*

-si crea una matrice k-1xk-1 di tutti 0-

*matH=j(&k-1,&k-1,0);*

-si inizializzano due cicli for-

```

do i=1 to &k-1;
do j=i+1 to &k;
-si calcola la distanza in valore assoluto -1 tra gli elementi del vettore x-
math[i,j-1]=abs(x[i]-x[j])-1;
-chiusura primo ciclo-
end;
-chiusura secondo ciclo-
end;
-si crea il data set H contenente la matrice H -
create H from math;
append from math;
quit;

```

Dopo aver ricavato le matrici K, L e H si calcola la statistica test:

$$T(\{rk(x_1)^{oss}, \dots, rk(x_n)^{oss}\}) = \sum_{1 \leq i \leq j \leq n} [ K(ij) + k(ij) * L(ij) + k(ij) * H(ij) ]$$

```

proc iml;
-si utilizza il data set Kp definendolo x-
use Kp;
read all var _num_ into x;
-si utilizza il data set Lp chiamandolo y-
use Lp;
read all var _num_ into y;
-si utilizza il data set H chiamandolo z-
use H;
read all var _num_ into z;
-si creano due matrici k-1xk-1 di tutti 0-
mat1=j(&k-1,&k-1,0);
mat2=j(&k-1,&k-1,0);
-si creano due vettori di tutti 1 per poter operare la somma degli elementi all'interno di
una matrice-

```

*v2=j(&k-1,1,1);*

*v1=j(1,&k-1,1);*

-si inizializzano due cicli for-

*do i=1 to &k-1;*

*do j=1 to &k-1;*

-si calcola il prodotto dei singoli elementi di x e y e di x e z ed i risultati vengono inseriti nelle matrici create precedentemente-

*mat1[i,j]=x[i,j]\*y[i,j];*

*mat2[i,j]=x[i,j]\*z[i,j];*

-si chiude il primo ciclo-

*end;*

-si chiude il secondo ciclo-

*end;*

-calcolo della statistica test-

*T2=v1\*x\*v2+v1\*mat1\*v2+v1\*mat2\*v2;*

-si crea il data set Tp2 -

*create Tp2 from T2;*

*append from T2;*

*quit;*

*run;*

-si crea un nuovo data set per avere Tp come titolo della col1-

*data Tp1;*

*set Tp2;*

*Tp=col1;*

*run;*

-si crea il data set Tp formato dalla sola colonna Tp essendo Tp1 formato da 2 colonne uguali (la prima chiamata col1, la seconda chiamata Tp)-

*data Tp;*

*set Tp1(keep=Tp);*

*run;*

Infine, si calcola l'I<sub>PG</sub>:

$$I_{PG} = \frac{T}{\frac{k * (k^2 - 1)}{3} - \frac{k * (k - 1)}{2}}$$

*proc iml;*

-si utilizza il data set Tp chiamandolo x-

*use Tp;*

*read all var \_num\_ into x;*

-si calcola la prima parte del denominatore-

*den=(&k\*(&k\*&k-1))/3;*

-si calcola la seconda parte del denominatore-

*deno=den-(&k\*(&k-1)/2);*

-calcolo dell'I<sub>PG</sub>-

*max1=x/deno;*

-si crea il data set maxp contenente l'indicatore I<sub>PG</sub>-

*create maxp from max1;*

*append from max1;*

*quit;*

#### 4.7.7 Indicatore I<sub>PG</sub> per la media

Si applica la statistica test al vettore della media di riferimento (&u) ed al vettore ottenuto dalla media aritmetica (rankm):

$$\mathbf{T}(\{\text{rk}(x_1)^{\text{OSS}}, \dots, \text{rk}(x_n)^{\text{OSS}}\}) = \sum_{1 \leq i \leq j \leq n} [ K(ij) + k(ij) * L(ij) + k(ij) * H(ij) ]$$

-si calcola la matrice K-

K (ij) assume valore 1 quando  $(\text{rk}(x_i)^{\text{rif}} > \text{rk}(x_j)^{\text{rif}})$  e  $(\text{rk}(x_i)^{\text{OSS}} < \text{rk}(x_j)^{\text{OSS}})$

o  $(rk(x_i)^{rif} < rk(x_j)^{rif})$  e  $(rk(x_i)^{oss} > rk(x_j)^{oss})$ ,

0 altrimenti.

*proc iml;*

-si utilizza il vettore ottenuto dalla media aritmetica ( rankm) chiamandolo x-  
*use rankm;*

*read all var \_num\_ into x;*

-si crea una matrice k-1xk-1 di tutti 0-

*matm=j(&k-1,&k-1,0);*

-si inizializzano due cicli for-

*do i=1 to &k-1;*

*do j=i+1 to &k;*

-se l'elemento i del vettore x è > dell'elemento j del vettore x si inserisce 1 nella matrice  
matm nel posto [i,j-1]-

*if x[i]>x[j]*

*then matm[i,j-1]=1;*

-si chiude il primo ciclo-

*end;*

-si chiude il secondo ciclo-

*end;*

-si crea il data set Km contenente la matrice K-

*create Km from matm;*

*append from matm;*

*quit;*

*run;*

-si calcola la matrice L-

$$L(ij) = | rk(x_i)^{rif} - rk(x_j)^{rif} | - 1$$

*proc iml;*

-si richiama il vettore della media di riferimento chiamandolo x-

```
use &u;
```

read all var *\_num\_* into x;

-si crea una matrice k-1xk-1 di tutti 0-

```
matlm=j(&k-1,&k-1,0);
```

-si inizializzano due cicli for-

```
do i=1 to &k-1;
```

```
do j=i+1 to &k;
```

-si calcola la distanza in valore assoluto -1 tra gli elementi del vettore x-

```
matlm[i,j-1]=abs(x[i]-x[j])-1;
```

-si chiude il primo ciclo-

```
end;
```

-si chiude il secondo ciclo-

```
end;
```

-si crea il data set Lm contenente la matrice L-

```
create Lm from matlm;
```

```
append from matlm;
```

```
quit;
```

```
run;
```

-si calcola la matrice H-

$$H(ij) = |rk(x_i)^{oss} - rk(x_j)^{oss}| - 1$$

```
proc iml;
```

-si utilizza il data set rankm chiamandolo x-

```
use rankm;
```

```
read all var _num_ into x;
```

-si crea una matrice k-1xk-1 di tutti 0-

```
matH=j(&k-1,&k-1,0);
```

-si inizializzano due cicli for-

```
do i=1 to &k-1;
```

```

do j=i+1 to &k;
-si calcola la distanza in valore assoluto -1 tra gli elementi del vettore x-
math[i,j-1]=abs(x[i]-x[j])-1;
-si chiude il primo ciclo-
end;
-si chiude il secondo ciclo-
end;
-si crea il data set H che contiene la matrice H-
create H from math;
append from math;
quit;
run;

```

Dopo aver ricavato le tre matrici si calcola la statistica test:

$$\mathbf{T}(\{\text{rk}(x_1)^{\text{OSS}}, \dots, \text{rk}(x_n)^{\text{OSS}}\}) = \sum_{1 \leq i \leq j \leq n} [ K(ij) + k(ij) * L(ij) + k(ij) * H(ij) ]$$

```

proc iml;
-si utilizza il data set Km definendolo x-
use Km;
read all var _num_ into x;
-si utilizza il data set Lm definendolo y-
use Lm;
read all var _num_ into y;
-si utilizza il data set H definendolo z-
use H;
read all var _num_ into z;
-si creano due matrici k-1xk-1 di tutti 0-
mat1=j(&k-1,&k-1,0);
mat2=j(&k-1,&k-1,0);
-si creano due vettori di tutti 1 per poter operare la somma degli elementi all'interno di
una matrice-

```

*v2=j(&k-1,1,1);*

*v1=j(1,&k-1,1);*

-si inizializzano due cicli for-

*do i=1 to &k-1;*

*do j=1 to &k-1;*

-si calcola il prodotto dei singoli elementi di x e y e di x e z e si inserisce nelle matrici create precedentemente-

*mat1[i,j]=x[i,j]\*y[i,j];*

*mat2[i,j]=x[i,j]\*z[i,j];*

-si chiude il primo ciclo-

*end;*

-si chiude il secondo ciclo-

*end;*

-calcolo della statistica test-

*T2=v1\*x\*v2+v1\*mat1\*v2+v1\*mat2\*v2;*

-si crea il data set Tm2 -

*create Tm2 from T2;*

*append from T2;*

*quit;*

*run;*

-si crea un nuovo data set per avere Tm come titolo della coll-

*data Tm1;*

*set Tm2;*

*Tm=coll;*

*run;*

-si crea il data set Tm formato dalla sola colonna Tm essendo Tm1 a sua volta formato da 2 colonne uguali (la prima chiamata coll, la seconda chiamata Tm)-

*data Tm;*

*set Tm1(keep=Tm);*

*run;*

Infine, si calcola l'IPG:

$$I_{PG} = \frac{T}{\frac{k * (k^2 - 1)}{3} - \frac{k * (k - 1)}{2}}$$

```

proc iml;
-si utilizza il data set Tm chiamandolo x-
use Tm;
read all var _num_ into x;
-si calcola la prima parte del denominatore-
den=(&k*(&k*&k-1))/3;
-si calcola la seconda parte del denominatore-
deno=den-(&k*(&k-1)/2);
-calcolo dell'IPG-
max1=x/deno;
-si crea il data set maxm contenente l'indicatore IPG-
create maxm from max1;
append from max1;
quit;
run;

```

#### 4.7.8 Confronto dei risultati

Si confrontano i risultati ottenuti dall'indice di Spearman  $I_S$  e dall'indicatore  $I_{PG}$ :

```

proc iml;
-si utilizza il data set maxm e lo si definisce Tm-
use maxm;
read all var _num_ into Tm;
-si utilizza il data set maxp e lo si definisce Tp-
use maxp;

```

```

read all var _num_ into Tp;
-si utilizza il data set Spearmanm e lo si definisce Sm-
use Spearmanm;
read all var _num_ into Sm;
-si utilizza il data set Spearmanp e lo si definisce Sp-
use Spearmanp;
read all var _num_ into Sp;
-si affiancano i quattro data set-
e=Tm||Tp||Sm||Sp;
-si crea il data set e1-
create e1 from e;
append from e;
quit;
run;
-si crea il data set fina dando i vari nomi alle colonne-
data fina;
set e1;
Tm=col1;
Tp=col2;
Sm=col3;
Sp=col4;
run;
-si crea il data set finale con le colonne rinominate-
data finale;
set fina(keep=pm pp sm sp);
proc print data=finale;
run;

-chiusura della macro-
%mend;

-richiamo della macro-

```

*%graduatoria(w=pesi,u=mediag,n=,k=);*

Valutando i risultati ottenuti dai due indici si può capire quale tra l’NPC ranking o la media aritmetica sia il migliore: più l’indice si avvicina al valore 1 più la classifica finale si allontana da quella reale.

#### **4.8 Clusters Analysis: implementazione dell’algoritmo di Zani per le classificazioni “sfocate”.**

Nell’algoritmo precedentemente proposto si è ipotizzato un campione omogeneo di intervistati. Se così non fosse risulterebbe necessario implementare un metodo che verifichi l’omogeneità delle persone intervistate. In questo paragrafo si presenta una particolare tecnica di Clusters Analysis definita “sfocata” (Zani 1989), basata sulla teoria dei “fuzzy sets” (Zadeh 1977; Bezdek 1981; Winham 1983). In essi l’appartenenza di un elemento ad un cluster è definita da un numero compreso tra 0 e 1: due elementi sono tanto più simili quanto più prossimo ad uno è il valore della loro funzione di appartenenza allo stesso cluster. Questo metodo permette, inoltre, di ottenere una classificazione congiunta delle  $n$  unità studiate in base ai  $k$  caratteri quantitativi e/o qualitativi. Proponiamo per ciascuno di tali fenomeni di individuare un’opportuna partizione delle  $n$  unità, con  $g_s$  gruppi ( $s=1, \dots, k$  e  $2 \leq g \leq n-1$ ). Per i caratteri qualitativi la partizione potrebbe essere semplicemente quella generata dalla suddivisione delle unità in base alle diverse modalità o categorie. Per i fenomeni quantitativi la ricerca della partizione più appropriata può essere ricondotta al problema della determinazione del numero e dell’ampiezza delle classi di una variabile unidimensionale, come ad esempio i quartili. Si definisce “grado di appartenenza congiunto” tra due unità statistiche  $a_i$  e  $a_j$ , la frequenza relativa delle  $k$  partizioni individuate in cui esse sono incluse in uno stesso gruppo:

$$z_{ij} = \frac{1}{k} \sum_{s=1}^k \delta_{(ij)s}$$

dove:  $\delta_{(ij)s} = \begin{cases} 1 & \text{se } a_i \text{ e } a_j \text{ sono nello stesso gruppo nella partizione } s - \text{esima.} \\ 0 & \text{altrimenti.} \end{cases}$

L'indice  $z_{ij}$  appartiene all'intervallo  $[0,1]$  ed assume il valore 1 quando c'è "piena appartenenza" delle due unità allo stesso gruppo in ciascuna delle  $k$  partizioni. In questo modo si ottiene una matrice di gradi di appartenenza, che costituisce il punto di partenza per calcolare le classificazioni sfocate delle  $n$  unità statistiche. Le fasi per

l'identificazione dei diversi clusters sono:

- a) si riuniscono tra loro le (eventuali) coppie di elementi con grado di appartenenza uguale ad 1, ottenendo in questo modo una partizione propria.
- b) nella matrice degli indici di similarità si cercano i valori del grado di appartenenza  $\alpha$  uguale a  $(k-1) / k$ . Se entrambe le unità che presentano tale valore non sono state assegnate ad un gruppo precedente, esse vengono riunite in un unico gruppo, con grado di appartenenza uguale ad  $\alpha$ . Se invece una delle unità era già stata inserita in un gruppo, l'altra viene aggregata al gruppo esistente, con grado di appartenenza al gruppo medesimo uguale ad  $\alpha$ .
- c) si itera la fase precedente considerando livelli via via decrescenti del grado di appartenenza:  $(k-2) / k$ ;  $(k-3) / k$ ; etc. In questi passi successivi si può manifestare anche il caso di coppie di unità che presentano tra loro il grado di appartenenza, considerato a quel livello, ma sono già state inserite in gruppi diversi. In questa circostanza suggeriamo di assegnare "in parte" ciascuna unità all'altro gruppo, con grado di appartenenza uguale al valore di  $\alpha$  in oggetto, purchè la somma per riga dei gradi risulti uguale ad 1.

#### **4.8.1 Implementazione in SAS**

Vengono ora presentati i comandi in SAS della Cluster Analysis, proponendo un esempio con 26 unità statistiche (comuni) e 6 caratteri quantitativi (varianza, densità, vecchiaia, giovanile, senile e stranieri). Questo algoritmo si ferma al calcolo della matrice dei gradi di appartenenza.

-Importazione dei dati-

```

libname var 'C:\percorso';
data var;
infile 'C:\percorso \varianza.txt';
input v1;
run;

libname dens 'C:\percorso';
data dens;
infile 'C:\percorso \densita.txt';
input d1;
run;

libname vec "C:\percorso";
data vec;
infile 'C:\percorso \vecchiaia.txt';
input ve1;
run;

libname giov 'C:\percorso';
data giov;
infile 'C:\percorso \giovanile.txt';
input g1;
run;

libname sen "C:\percorso";
data sen;
infile 'C:\percorso \senile.txt';
input s1;
run;

libname stran 'C:\percorso';
data stran;
infile 'C:\percorso \stranieri.txt';
input st1;
run;

-si utilizza l'ambiente iml-
proc iml;

```

-si utilizzano i vari data set definendoli come variabili-

```
use var;
```

```
read all var _num_ into x;
```

```
use dens;
```

```
read all var _num_ into y;
```

```
use vec;
```

```
read all var _num_ into z;
```

```
use giov;
```

```
read all var _num_ into a;
```

```
use sen;
```

```
read all var _num_ into b;
```

```
use stran;
```

```
read all var _num_ into c;
```

-si costruisce un nuovo data set che ha per colonne le 6 caratteristiche-

```
d=x||y||z||a||b||c;
```

```
create prova from d;
```

```
append from d;
```

```
quit;
```

-si stampa il nuovo data set prova-

```
proc print data=prova;
```

```
run;
```

```
data prova;
```

```
set prova;
```

```
a=1;
```

```
run;
```

```
proc sort;
```

```
by a;
```

```
run;
```

-si utilizza l'ambiente macro-

```
%macro uni;
```

```
%let lista=col1 col2 col3 col4 col5 col6;
```

-si inizializza un ciclo for-

```

%do i=1 %to 6;
-calcolo dei quartili-
proc univariate data=prova noprint;
var %scan(&lista,&i);
output out=primo&i MEDIAN=med_&i Q1=Q1_&i Q3=Q3_&i;
run;
-si stampano i vari data set-
proc print data=primo&i;
run;
-chiusura del ciclo for-
%end;
si costruisce un data set con i risultati ottenuti precedentemente, ponendoli uno sopra
l'altro-
data prova2;
merge %do i=1 %to 6; primo&i %end;;
a=1;
run;
proc sort;
by a;
run;
-si uniscono in un unico data set i data set prova e prova2-
data prova3;
merge prova prova2;
by a;
run;
-si dividono i dati in quattro gruppi-
data prova4;
set prova3;
%do i=1 %to 6;
if %scan(&lista,&i)<=Q1_&i then gruppo_&i=1;
else if %scan(&lista,&i)<=med_&i then gruppo_&i=2;
else if %scan(&lista,&i)<=Q3_&i then gruppo_&i=3;

```

```

else if %scan(&lista,&i)>Q3_&i then gruppo_&i=4;
%end;

run;

-si traspone il data set prova4 ottenendo il data set prova5-
proc transpose data=prova4 out=prova5;
run;

-si definiscono le colonne del data set prova5-
proc print data=prova5;
where_NAME_
in("gruppo_1","gruppo_2","gruppo_3","gruppo_4","gruppo_5","gruppo_6");
run;

-si analizzano i diversi gruppi-
data prova6;
set prova5;
where_NAME_
in("gruppo_1","gruppo_2","gruppo_3","gruppo_4","gruppo_5","gruppo_6");
%do i=1 %to 26;
    %do j=1 %to 26;
        if col&i=col&j then do;
            d_&i._&j=1;
            end;
        else do;
            d_&i._&j=0;
            end;
        %end;
    %end;
%end;
run;

-si traspone il data set prova6 ottenendo prova7-
proc transpose data=prova6 out=prova7;
run;

```

-si calcolano I gradi di appartenenza  $z_i$ -

```
data prova8;
```

```
set prova7;
```

```
zeta=sum(gruppo_1, gruppo_2,gruppo_3,gruppo_4, gruppo_5, gruppo_6)/6;
```

```
run;
```

-si crea il data set prova9 con la sola colonna degli  $z_i$ -

```
data prova9;
```

```
set prova8(keep=zeta);
```

```
proc print data=prova9;
```

```
run;
```

-costruzione della matrice dei gradi di appartenenza-

```
proc iml;
```

```
use prova9(firstobs=27 obs=702);
```

```
read all var _num_ into x;
```

```
mat=j(26,26,0);
```

```
do i=1 to 676;
```

```
mat[i]=x[i];
```

```
end;
```

```
create mat1 from mat;
```

```
append from mat;
```

```
quit;
```

```
proc iml;
```

```
use mat1;
```

```
read all var _num_ into x;
```

```
mat=j(26,26,0);
```

```
do i=1 to 25;
```

```
do j=i+1 to 26;
```

```
mat[i,j]=x[i,j];
```

```
end;
```

```
end;
```

```
print mat;
```

```
create matf from mat;
```

```
append from mat;  
quit;
```



## *Capitolo 5*

# ***STUDIO COMPARATIVO DI SIMULAZIONE: NPC RANKING VS MEDIA***

### **5.1 Introduzione**

Nel capitolo precedente si è presentato, nell'ambito della Conjoint Analysis, il metodo delle combinazioni non parametriche (NPC ranking) come alternativa alla media aritmetica, attualmente utilizzata in letteratura, per l'aggregazione dei giudizi di preferenza. In questo capitolo si intende confrontare le due tecniche precedentemente menzionate attraverso simulazioni di dati generati casualmente da diverse distribuzioni discrete e continue. Questo al fine di verificare se e quando l'NPC ranking risulti migliore della media. Inoltre, dopo aver calcolato tutte le possibili permutazioni di un ipotetico vettore di riferimento, si sono analizzati il test di Spearman TS ed il test di permutazione TPG per valutarne la potenza.

## 5.2 Cardinalità dello spazio delle permutazioni

Il problema del confronto tra l’NPC ranking e la media viene affrontato simulando un caso reale caratterizzato da tre attributi (A B C) su due livelli (+ e -):

A	B	C	graduatoria di riferimento
<b>Cart1</b>	-	-	1
<b>Cart2</b>	-	-	2
<b>Cart3</b>	-	+	3
<b>Cart4</b>	-	+	4
<b>Cart5</b>	+	-	5
<b>Cart6</b>	+	-	6
<b>Cart7</b>	+	+	7
<b>Cart8</b>	+	+	8

Tabella 5.1: graduatoria di riferimento

in cui 8 identifica il cartellino migliore. Il piano fattoriale in questo caso è  $2^3$  (quindi con un totale di 8 cartellini), tutti i possibili confronti a coppie sono  $\binom{8}{2} = 28$  mentre le permutazioni risultano:  $8! = 40320$  (come illustrato in tabella 5.2):

n°perm.	Cart1	Cart2	Cart3	Cart4	Cart5	Cart6	Cart7	Cart8
1	1	2	3	4	5	6	7	8
2	1	2	3	4	5	6	8	7
3	1	2	3	4	5	7	6	8
4	1	2	3	4	5	7	8	6
5	1	2	3	4	5	8	6	7
6	1	2	3	4	5	8	7	6
7	1	2	3	4	6	5	7	8
8	1	2	3	4	6	5	8	7
9	1	2	3	4	6	7	5	8
10	1	2	3	4	6	7	8	5
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
20156	4	8	7	6	5	1	3	2
20157	4	8	7	6	5	2	1	3
20158	4	8	7	6	5	2	3	1
20159	4	8	7	6	5	3	1	2
20160	4	8	7	6	5	3	2	1
20161	5	1	2	3	4	6	7	8
20162	5	1	2	3	4	6	8	7
20163	5	1	2	3	4	7	6	8
20164	5	1	2	3	4	7	8	6
20165	5	1	2	3	4	8	6	7
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
40311	8	7	6	5	3	2	1	4
40312	8	7	6	5	3	2	4	1
40313	8	7	6	5	3	4	1	2
40314	8	7	6	5	3	4	2	1
40315	8	7	6	5	4	1	2	3
40316	8	7	6	5	4	1	3	2
40317	8	7	6	5	4	2	1	3
40318	8	7	6	5	4	2	3	1
40319	8	7	6	5	4	3	1	2
40320	8	7	6	5	4	3	2	1

Tabella 5.2: esempi di permutazioni della graduatoria di riferimento

### 5.3 Descrizione dello studio di simulazione

In questa sessione si ipotizza di sottoporre a cinque persone (Risp\_j con j=1,...,5) gli otto cartellini sopra riportati (Tabella 5.1).

	Risp_1	Risp_2	Risp_3	Risp_4	Risp_5	C.f.NPC	C.f.media
<b>Cartellino 1</b>	$y_{1,1}$	$y_{1,2}$	$y_{1,3}$	$y_{1,4}$	$y_{1,5}$	$r(z)_1$	$\bar{y}_1$
<b>Cartellino 2</b>	$y_{2,1}$	$y_{2,2}$	$y_{2,3}$	$y_{2,4}$	$y_{2,5}$	$r(z)_2$	$\bar{y}_2$
<b>Cartellino 3</b>	$y_{3,1}$	$y_{3,2}$	$y_{3,3}$	$y_{3,4}$	$y_{3,5}$	$r(z)_3$	$\bar{y}_3$
<b>Cartellino 4</b>	$y_{4,1}$	$y_{4,2}$	$y_{4,3}$	$y_{4,4}$	$y_{4,5}$	$r(z)_4$	$\bar{y}_4$
<b>Cartellino 5</b>	$y_{5,1}$	$y_{5,2}$	$y_{5,3}$	$y_{5,4}$	$y_{5,5}$	$r(z)_5$	$\bar{y}_5$
<b>Cartellino 6</b>	$y_{6,1}$	$y_{6,2}$	$y_{6,3}$	$y_{6,4}$	$y_{6,5}$	$r(z)_6$	$\bar{y}_6$
<b>Cartellino 7</b>	$y_{7,1}$	$y_{7,2}$	$y_{7,3}$	$y_{7,4}$	$y_{7,5}$	$r(z)_7$	$\bar{y}_7$
<b>Cartellino 8</b>	$y_{8,1}$	$y_{8,2}$	$y_{8,3}$	$y_{8,4}$	$y_{8,5}$	$r(z)_8$	$\bar{y}_8$

Tabella 5.3: matrice delle risposte (8x5)

Gli  $y_{i,j}$  riportati nella tabella 5.3, con  $i = 1, \dots, 8$ , corrispondono alla valutazione della  $j$ -esima persona sull' $i$ -esimo cartellino, mentre  $r(z)_i$  rappresenta il valore espresso in rango dell' $i$ -esimo cartellino nella classifica finale calcolata con il metodo NPC ranking (C.f.NPC) e  $\bar{y}_i$  è il valore espresso in rango dell' $i$ -esimo cartellino nella classifica finale calcolata con la media (C.f.media), con  $i = 1, \dots, 8$  e  $j = 1, \dots, 5$ .

Gli ipotetici giudizi degli intervistati sono stati ottenuti grazie ad una generazione casuale di dati, sommando o sottraendo alla graduatoria di riferimento degli errori casuali calcolati da distribuzioni continue e discrete quali: Normale, Gamma, Uniforme e Binomiale. Per ogni distribuzione si sono calcolate mille matrici diverse come quella sopra indicata. Su ognuna di queste si è applicato l'algoritmo presentato nel quarto capitolo ottenendo così mille graduatorie finali sia con l'NPC ranking che con la media.

	<b>Class.fin.1</b>	<b>Class.fin.2</b>	<b>...</b>	<b>Class.fin.1000</b>
<b>Cartellino 1</b>	$r(z)_{1,1}$	$r(z)_{1,2}$	...	$r(z)_{1,1000}$
<b>Cartellino 2</b>	$r(z)_{2,1}$	$r(z)_{2,2}$	...	$r(z)_{2,1000}$
<b>Cartellino 3</b>	$r(z)_{3,1}$	$r(z)_{3,2}$	...	$r(z)_{3,1000}$
<b>Cartellino 4</b>	$r(z)_{4,1}$	$r(z)_{4,2}$	...	$r(z)_{4,1000}$
<b>Cartellino 5</b>	$r(z)_{5,1}$	$r(z)_{5,2}$	...	$r(z)_{5,1000}$
<b>Cartellino 6</b>	$r(z)_{6,1}$	$r(z)_{6,2}$	...	$r(z)_{6,1000}$
<b>Cartellino 7</b>	$r(z)_{7,1}$	$r(z)_{7,2}$	...	$r(z)_{7,1000}$
<b>Cartellino 8</b>	$r(z)_{8,1}$	$r(z)_{8,2}$	...	$r(z)_{8,1000}$

**Tabella 5.4: classifiche finali ottenute con l’NPC ranking**

Ogni  $r(z)_{i,j}$  presente nella tabella 5.4 corrisponde al valore espresso in rango dell’i-esimo cartellino nella j-esima classifica finale.

	<b>Class.fin.1</b>	<b>Class.fin.2</b>	<b>...</b>	<b>Class.fin.1000</b>
<b>Cartellino 1</b>	$\bar{y}_{1,1}$	$\bar{y}_{1,2}$	...	$\bar{y}_{1,1000}$
<b>Cartellino 2</b>	$\bar{y}_{2,1}$	$\bar{y}_{2,2}$	...	$\bar{y}_{2,1000}$
<b>Cartellino 3</b>	$\bar{y}_{3,1}$	$\bar{y}_{3,2}$	...	$\bar{y}_{3,1000}$
<b>Cartellino 4</b>	$\bar{y}_{4,1}$	$\bar{y}_{4,2}$	...	$\bar{y}_{4,1000}$
<b>Cartellino 5</b>	$\bar{y}_{5,1}$	$\bar{y}_{5,2}$	...	$\bar{y}_{5,1000}$
<b>Cartellino 6</b>	$\bar{y}_{6,1}$	$\bar{y}_{6,2}$	...	$\bar{y}_{6,1000}$
<b>Cartellino 7</b>	$\bar{y}_{7,1}$	$\bar{y}_{7,2}$	...	$\bar{y}_{7,1000}$
<b>Cartellino 8</b>	$\bar{y}_{8,1}$	$\bar{y}_{8,2}$	...	$\bar{y}_{8,1000}$

**Tabella 5.5: classifiche finali ottenute con la media aritmetica**

Ogni  $\bar{y}_{i,j}$  presente nella tabella 5.5 corrisponde al valore espresso in rango dell’i-esimo cartellino nella j-esima classifica finale.

Successivamente si sono analizzate le singole classifiche finali precedentemente calcolate, attraverso l’indicatore di permutazione sulle graduatorie  $I_{PG}$  e l’indice di Spearman  $I_S$ , al fine di verificare quanto differiscano dalla graduatoria di riferimento. Così facendo si ottengono mille indici compresi nell’intervallo  $[0,1]$  per entrambi i metodi di confronto ( $I_S$  e  $I_{PG}$ ), rispettivamente per l’NPC ranking e per la media

aritmetica: più il valore degli indici si avvicina a zero più la graduatoria osservata rispecchia quella di riferimento.

	NPC ranking		Media aritmetica	
	$I_S$	$I_{PG}$	$I_S$	$I_{PG}$
<b>Grad. fin. 1</b>	$i_{1,s}$	$i_{1,t}$	$i_{1,s}$	$i_{1,t}$
<b>Grad. fin. 2</b>	$i_{2,s}$	$i_{2,t}$	$i_{2,s}$	$i_{2,t}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
<b>Grad. fin. 1000</b>	$i_{1000,s}$	$i_{1000,t}$	$i_{1000,s}$	$i_{1000,t}$

Tabella 5.6: indici di confronto calcolati sulle graduatorie finali ottenute dalle rispettive matrici.

Dove:

$i_{k,s}$  è l'indice di Spearman  $I_S$  per la k-esima graduatoria finale.

$i_{k,t}$  è l'indicatore di permutazione delle graduatorie  $I_{PG}$  per la k-esima graduatoria finale.

Infine, avendo a disposizione tutti gli indici, si possono confrontare tra loro sia i risultati ottenuti con l'NPC ranking che con la media sia gli indicatori di confronto. In particolare:

1) confronto tra l'NPC ranking (p) e la media (m) attraverso l' $I_{PG}$  (T):

$$\frac{\#(T_{pi} \leq T_{mi})}{1000} \quad \frac{\#(T_{mi} \leq T_{pi})}{1000} \quad \text{dove } i=1, \dots, 1000.$$

Il risultato di ognuno dei due confronti assume valori nell'intervallo [0,1], mentre la loro somma è compresa nell'intervallo [1,2], poiché gli ex-aequo vengono contati in

entrambe le disuguaglianze. Ad esempio se  $\frac{\#(T_{pi} \leq T_{mi})}{1000} = 0.9$  e  $\frac{\#(T_{mi} \leq T_{pi})}{1000} = 0.3$ ,

l'NPC ranking è preferibile alla media, nel senso che le graduatorie calcolate con il primo metodo si avvicinano di più a quella di riferimento rispetto a quelle ottenute con la media.

2) confronto tra l'NPC ranking (p) e la media (m) attraverso l' $I_S$  (S)-

$$\frac{\#(S_{pi} \leq S_{mi})}{1000}$$

$$\frac{\#(S_{mi} \leq S_{pi})}{1000}$$

Il risultato di ognuno dei due confronti assume valori nell'intervallo [0,1], mentre la loro somma è compresa nell'intervallo [1,2], poiché gli ex-aequo vengono contati in entrambe le disuguaglianze. Ad esempio se  $\frac{\#(S_{pi} \leq S_{mi})}{1000} = 0.45$  e  $\frac{\#(S_{mi} \leq S_{pi})}{1000} = 0.6$ ,

la media è preferibile all'NPC ranking, nel senso che le graduatorie calcolate con la media si avvicinano di più a quella di riferimento rispetto a quelle ottenute con l'NPC ranking.

3) confronto tra l'I<sub>PG</sub> (T) e l'I<sub>S</sub> (S) entrambi calcolati sulle classifiche finali ottenute con l'NPC ranking (p)-

$$\frac{\#(T_{pi} \leq S_{pi})}{1000}$$

$$\frac{\#(S_{pi} \leq T_{pi})}{1000}$$

Il risultato di ognuno dei due confronti assume valori nell'intervallo [0,1], mentre la loro somma è compresa nell'intervallo [1,2], poiché gli ex-aequo vengono contati in entrambe le disuguaglianze. Ad esempio se  $\frac{\#(T_{pi} \leq S_{pi})}{1000} = 0.8$  e  $\frac{\#(S_{pi} \leq T_{pi})}{1000} = 0.3$ ,

l'I<sub>PG</sub> è preferito all'I<sub>S</sub> per analizzare le classifiche ottenute con l'NPC ranking.

4) confronto tra l'I<sub>PG</sub> (T) e l'I<sub>S</sub> (S) entrambi calcolati sulle classifiche finali ottenute con la media (m)-

$$\frac{\#(T_{mi} \leq S_{mi})}{1000}$$

$$\frac{\#(S_{mi} \leq T_{mi})}{1000}$$

Il risultato di ognuno dei due confronti assume valori nell'intervallo [0,1], mentre la loro somma è compresa nell'intervallo [1,2], poiché gli ex-aequo vengono contati in entrambe le disuguaglianze. Ad esempio se  $\frac{\#(T_{mi} \leq S_{mi})}{1000} = 0.9$  e  $\frac{\#(S_{mi} \leq T_{mi})}{1000} = 0.2$ ,

l'I<sub>PG</sub> è preferito all'I<sub>S</sub> per analizzare le classifiche ottenute con la media.

## 5.4 Studio di simulazioni

Gli ipotetici giudizi degli intervistati  $y_{i,j}$  sono stati ottenuti grazie ad una generazione casuale di dati, sommando o sottraendo alla graduatoria di riferimento degli errori casuali calcolati da alcune distribuzioni discrete e continue, formalmente:

$$y_{i,j} = \mu_i + \varepsilon_{i,j}$$

dove  $\mu_i$  è il valore dell' $i$ -esimo cartellino nella graduatoria di riferimento e  $\varepsilon_{i,j}$  è l'errore generato dalle diverse distribuzioni riferito all' $i$ -esimo cartellino della  $j$ -esima persona. I giudizi  $y_{ij}$  vengono poi ordinati per ranghi, seguendo la regola grande è meglio. Le classifiche finali ottenute con l'NPC ranking e con la media possono differire da quella di riferimento sia per la distanza tra le posizioni sia per il numero di posizioni scambiate. Si definisce "distanza tra le posizioni", di un'ipotetica classifica con  $k$  elementi e con  $i = 1, \dots, k$ , la differenza in valore assoluto tra l' $i$ -esimo elemento di questa e l' $i$ -esimo elemento della classifica di riferimento, anch'essa formata da  $k$  elementi ed  $i = 1, \dots, k$ . I valori di questa differenza appartengono all'intervallo  $[0, k-1]$ , dove 0 indica che non c'è distanza tra due elementi mentre  $k-1$  rappresenta la distanza massima. Ad esempio, se si ha come classifica di riferimento (1 2 3 4) ed un'ipotetica classifica osservata (3 1 2 4), si ottengono i seguenti risultati:

<b>Class. di Riferimento</b>	<b>Class. Osservata</b>	<b>Distanza tra posizioni</b>
<b>1</b>	<b>3</b>	<b>2</b>
<b>2</b>	<b>1</b>	<b>1</b>
<b>3</b>	<b>2</b>	<b>1</b>
<b>4</b>	<b>4</b>	<b>0</b>

Si definisce "numero di posizione scambiate" il numero di volte in cui la distanza tra le posizioni risulta diversa da zero come nella seguente tabella:

Class. di Riferimento	Class. Osservata	Distanza tra posizioni	Posizioni Scambiate
1	3	2	1
2	1	1	1
3	2	1	1
4	4	0	0
			<b>N° di pos. Scamb. = 3</b>

### 5.4.1 Comportamento con errori di tipo Normale

Una variabile casuale continua  $\varepsilon$  con funzione di densità in  $t$

$$\frac{1}{\sqrt{2\pi\sigma}} \exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right\} \quad \text{con } -\infty < t < +\infty$$

con  $\sigma > 0$ , si definisce normale con parametri  $\mu$  e  $\sigma^2$ , per brevità indicata con

$$\varepsilon \sim N(\mu, \sigma^2).$$

Per calcolare l'algoritmo in SAS si sono considerati i seguenti valori:  $\mu = 0$  e  $\sigma^2 = 1$ ,

```
ncas[i, j]=sqrt(1)*normal(&r)+rel[i];
```

dove "rel[i]" è il vettore della graduatoria di riferimento ed "&r" è la componente casuale "seed" e  $\text{sqrt}(1) = \sigma$ .

I risultati ottenuti dall'analisi sono riassunti nella seguente tabella:

$$1) \frac{\#(T_p \leq T_m)}{1000} = 0.531$$

$$\frac{\#(T_m \leq T_p)}{1000} = 0.771$$

$$2) \frac{\#(S_p \leq S_m)}{1000} = 0.526$$

$$\frac{\#(S_m \leq S_p)}{1000} = 0.772$$

$$3) \frac{\#(T_p \leq S_p)}{1000} = 1$$

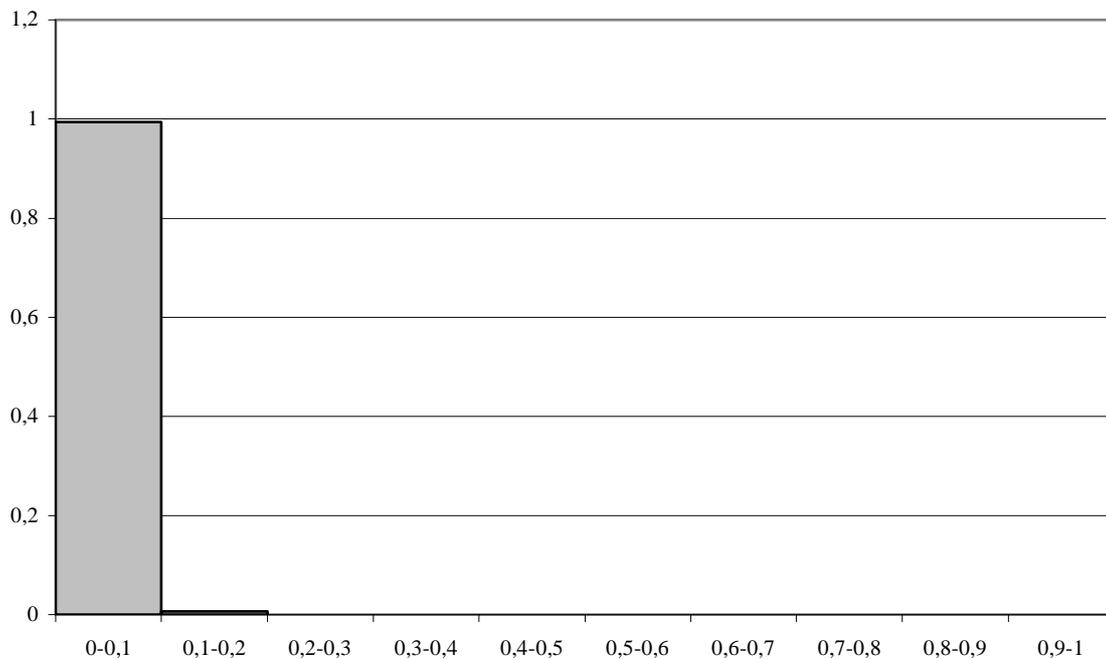
$$\frac{\#(S_p \leq T_p)}{1000} = 0.076$$

$$4) \frac{\#(T_m \leq S_m)}{1000} = 1$$

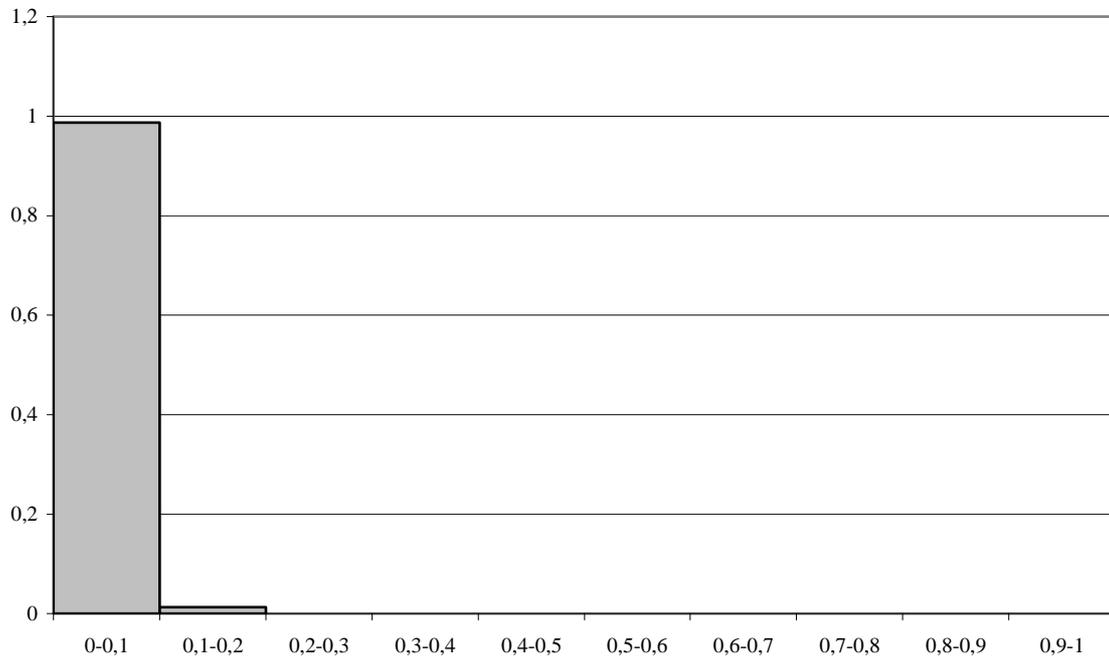
$$\frac{\#(S_m \leq T_m)}{1000} = 0.125$$

Osservando gli indici, si può notare che nel confronto (1) l'NPC ranking è leggermente peggiore della media; questa tesi è rafforzata anche dai risultati ottenuti nel confronto (2) con l' $I_S$ . Mentre dall'analisi dei due indici di confronto emerge che l' $I_{PG}$  è molto più potente dell'indice  $I_S$ . Inoltre si sono analizzate le classifiche finali con dei grafici.

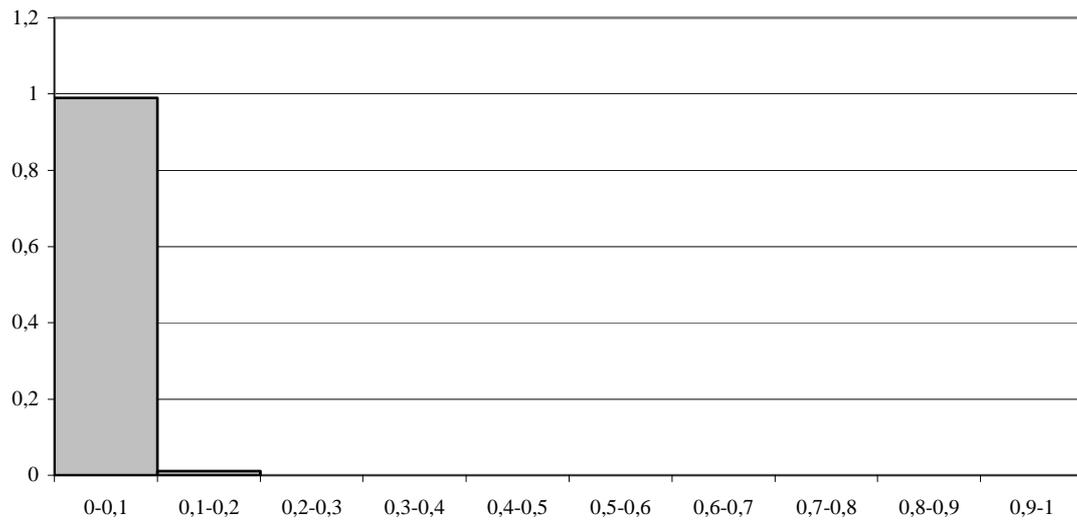
### **$I_{PG}$ (errori di tipo Normale, media)**



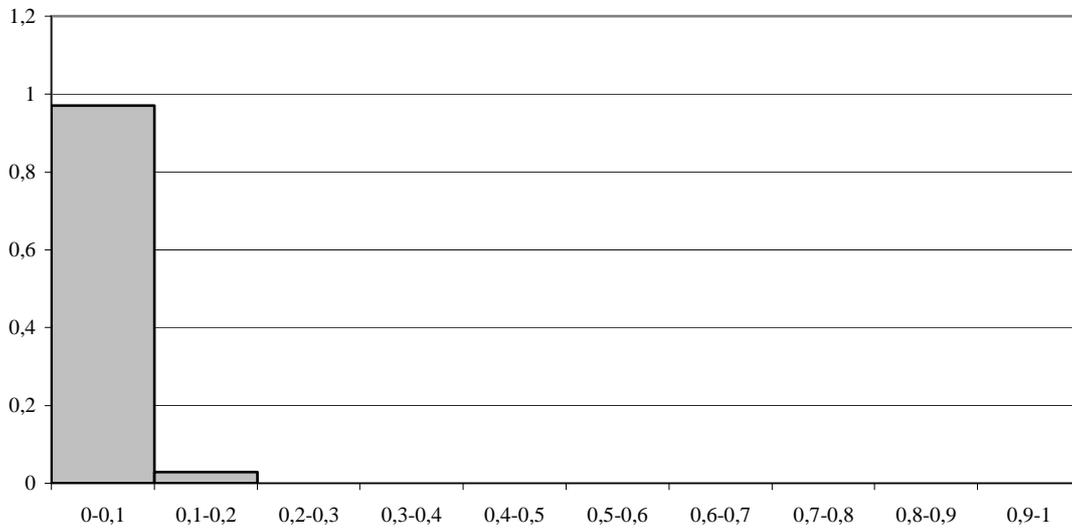
### IPG (errori di tipo Normale, NPC ranking)



### Is (errori di tipo Normale, media)



### Is (errori di tipo Normale, NPC ranking)



I grafici sopra riportati rappresentano la suddivisione in percentuale delle graduatorie finali ottenute con la media e con l’NPC ranking (**asse delle ordinate**), in base ai valori dell’ $I_{PG}$  e dell’indice  $I_S$  calcolati su di esse (**asse delle ascisse**). In tutti i grafici si può notare come la maggior parte degli indici  $I_S$  e  $I_{PG}$  si distribuisca nell’intervallo [0-0.1]. Questo significa che gli errori generati dalla distribuzione Normale hanno modificato di poco i giudizi simulati su cui si è calcolato l’algoritmo e di conseguenza le classifiche finali ottenute risultano molto simili a quella di riferimento. In questo contesto l’utilizzo della media risulta una buona soluzione per l’aggregazione dei giudizi di preferenza.

#### 5.4.2 Comportamento con errori di tipo Gamma

Una variabile casuale continua  $\varepsilon$  è di tipo gamma di indice  $\omega$  e parametro di scala  $\lambda$  se ha funzione di densità in  $t$

$$f(t; \omega, \lambda) = \begin{cases} \lambda^\omega t^{\omega-1} e^{-\lambda t} / \Gamma(\omega) & \text{per } t > 0, \\ 0 & \text{per } t \leq 0, \end{cases}$$

dove  $\Gamma(\omega) \approx \sqrt{2\pi} \omega^{-\frac{1}{2}} e^{-\omega}$  e per qualche  $\omega$  e  $\lambda$  positivi;  $\varepsilon \sim G(\omega, \lambda)$ .

Per calcolare l'algorithmo in SAS si è considerato  $\omega = 1$ :

```
ncas[i,j]= x[i]+rangam(&r,1)
```

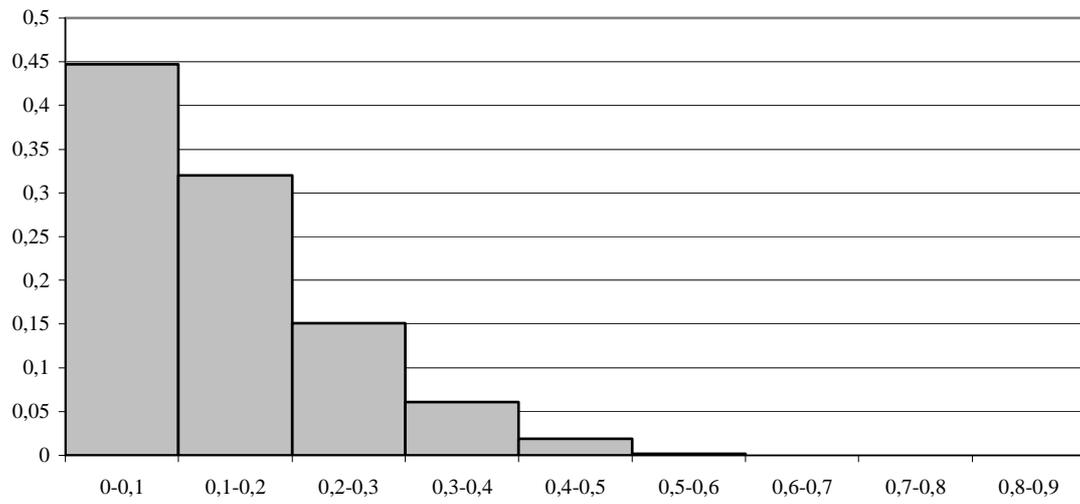
dove "x[i]" è il vettore della graduatoria di riferimento ed "&r" è la componente casuale "seed" e  $\omega = 1$ .

I risultati ottenuti dall'analisi sono riassunti nella seguente tabella:

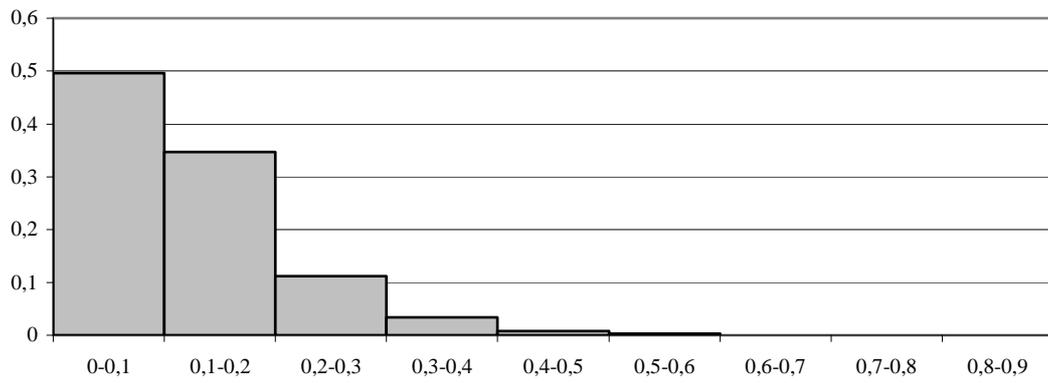
1) $\frac{\#(T_p \leq T_m)}{1000} = 0.65$	$\frac{\#(T_m \leq T_p)}{1000} = 0.447$
2) $\frac{\#(S_p \leq S_m)}{1000} = 0.592$	$\frac{\#(S_m \leq S_p)}{1000} = 0.461$
3) $\frac{\#(T_p \leq S_p)}{1000} = 0.996$	$\frac{\#(S_p \leq T_p)}{1000} = 0.015$
4) $\frac{\#(T_m \leq S_m)}{1000} = 0.991$	$\frac{\#(S_m \leq T_m)}{1000} = 0.021$

A differenza della distribuzione Normale, precedentemente analizzata, si può notare che nel confronto (1) l'NPC ranking risulti migliore della media, inoltre questa tesi è rafforzata anche dai risultati ottenuti nel confronto (2) con l'indice di Spearman ( $I_S$ ). Dall'analisi dei due indici di confronto emerge invece che l' $I_{PG}$  è più potente dell' $I_S$ .

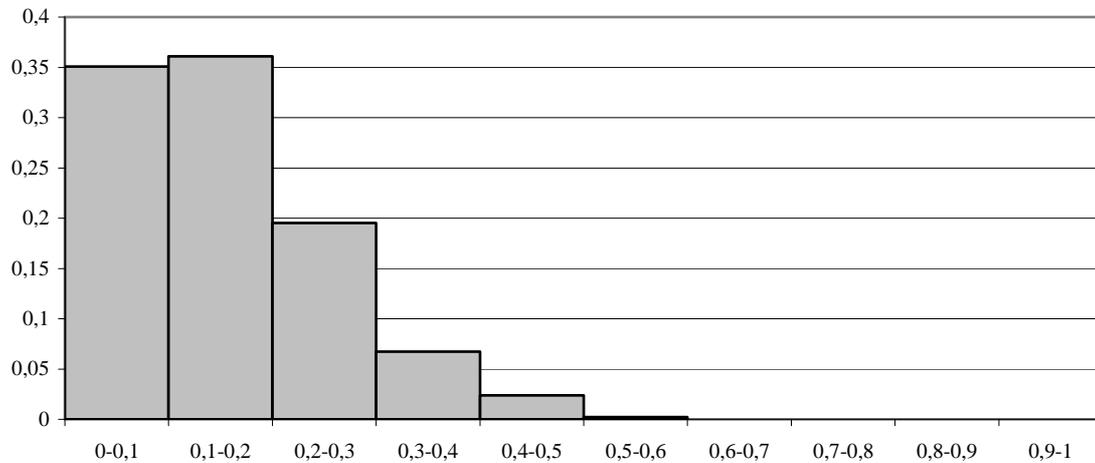
**IPG (errori di tipo Gamma, media)**



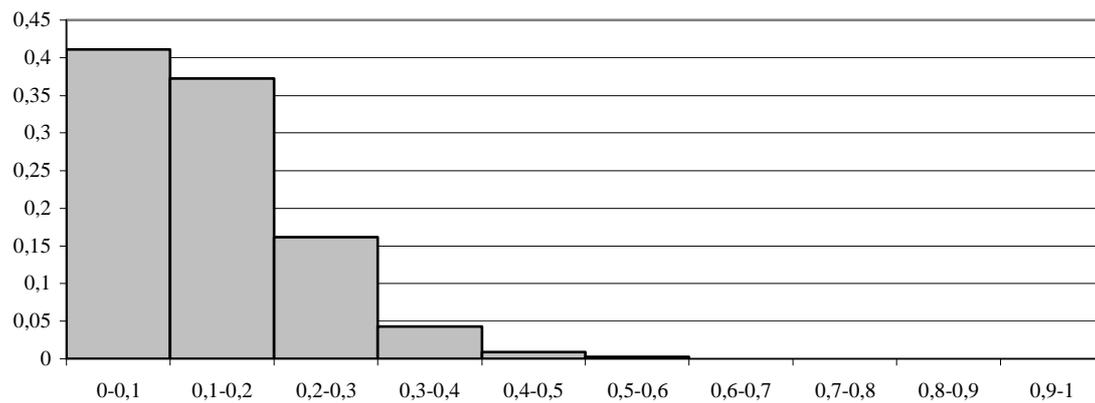
**IPG (errori di tipo Gamma, NPC ranking)**



**Is (errori di tipo Gamma, media)**



**Is (errori di tipo Gamma, NPC ranking)**



I grafici sopra riportati rappresentano la suddivisione in percentuale delle graduatorie finali ottenute con la media e con l’NPC ranking (**asse delle ordinate**), in base ai valori dell’indicatore  $I_{PG}$  e dell’indice  $I_S$  calcolati su di esse (**asse delle ascisse**). Nei due grafici in cui le classifiche sono state ottenute con l’NPC ranking si può notare come la percentuale degli indici  $I_S$  e  $I_{PG}$  negli intervalli [0-0,1] e [0,1-0,2] risulta più elevata rispetto al caso della media. Mentre negli intervalli [0,2-0,3], [0,3-0,4] e [0,4-0,5] ci sono più indici  $I_S$  e  $I_{PG}$  nei grafici in cui le classifiche sono state ottenute con la media. Questo significa che gli errori generati dalla distribuzione Gamma hanno modificato, in alcuni casi di molto in altri di poco, i giudizi simulati su cui si è calcolato l’algoritmo e

di conseguenza le classifiche finali risultano in certi casi distanti ed in altri sufficientemente vicine a quella di riferimento. In questo contesto l'utilizzo dell'NPC ranking per l'aggregazione dei giudizi di preferenza può essere considerato complessivamente migliore della media.

### 5.4.3 Comportamento con errori di tipo Uniforme

Una variabile casuale continua  $\varepsilon$  con funzione di densità in  $t$

$$f(t; a, b) = \begin{cases} \frac{1}{b-a} & \text{per } t \in (a, b), \\ 0 & \text{altrimenti,} \end{cases}$$

è detta uniforme in  $(a, b)$ , e si scrive  $\varepsilon \sim U(a, b)$ , posto che  $a$  e  $b$  siano reali con  $a < b$ .

Per calcolare l'algoritmo in SAS si è presa in considerazione una distribuzione uniforme di parametri  $(0, 1)$ :

```
ncas[i, j]=rel[i]+ranuni(&r)
```

dove "rel[i]" è il vettore della graduatoria di riferimento ed "&r" è la componente casuale "seed".

I risultati ottenuti dall'analisi sono riassunti nella seguente tabella:

$$1) \frac{\#(T_p \leq T_m)}{1000} = 0.441$$

$$\frac{\#(T_m \leq T_p)}{1000} = 0.757$$

$$2) \frac{\#(S_p \leq S_m)}{1000} = 0.439$$

$$\frac{\#(S_m \leq S_p)}{1000} = 0.758$$

$$3) \frac{\#(T_p \leq S_p)}{1000} = 1$$

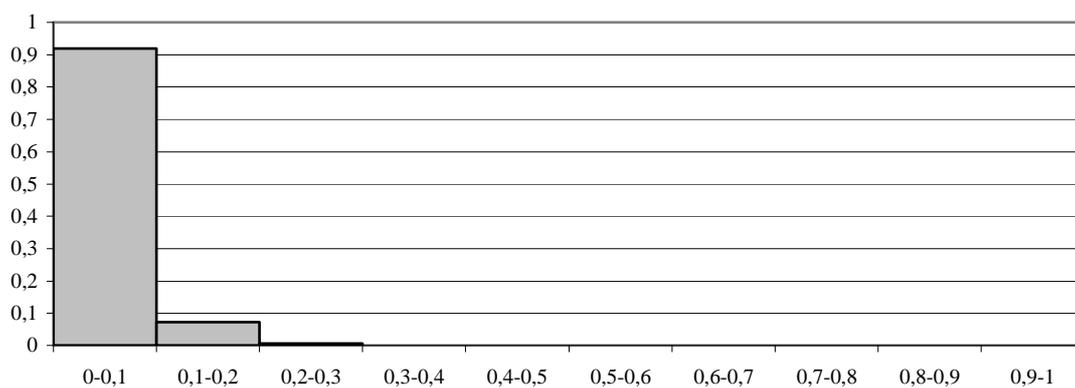
$$\frac{\#(S_p \leq T_p)}{1000} = 1$$

$$4) \frac{\#(T_m \leq S_m)}{1000} = 1$$

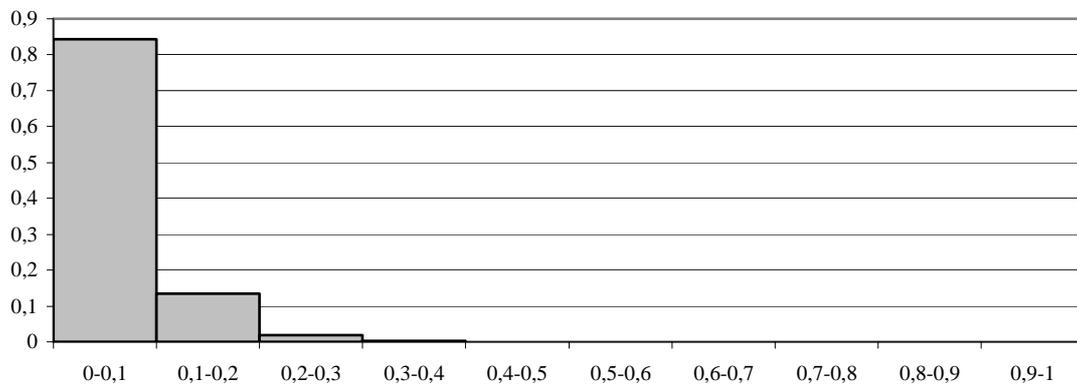
$$\frac{\#(S_m \leq T_m)}{1000} = 0.017$$

Come nel caso della distribuzione Normale, si può notare che nel confronto (1) l'NPC ranking è peggiore della media; questa tesi è rafforzata anche dai risultati ottenuti nel confronto (2) con l'indice  $I_S$ . Mentre dall'analisi dei due indici di confronto emerge che l' $I_{PG}$  è più potente dell'indice  $I_S$ .

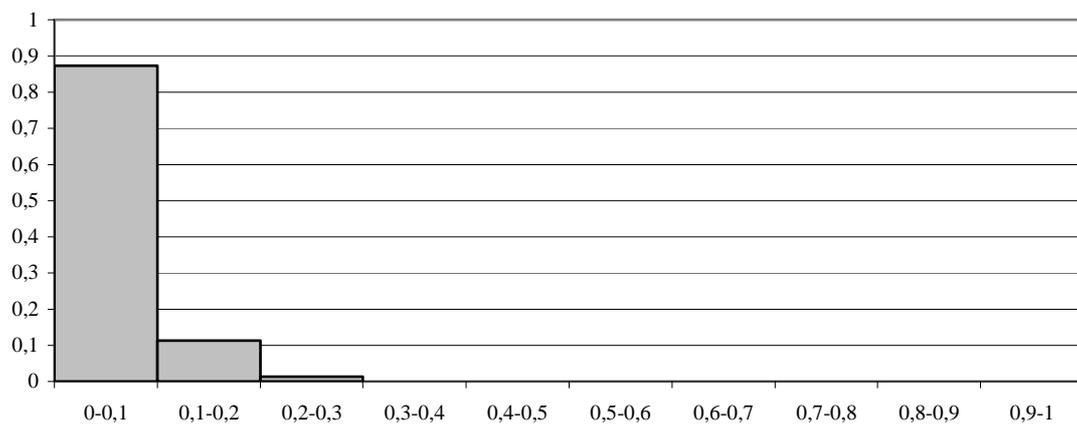
### **$I_{PG}$ (errori di tipo Uniforme, media)**



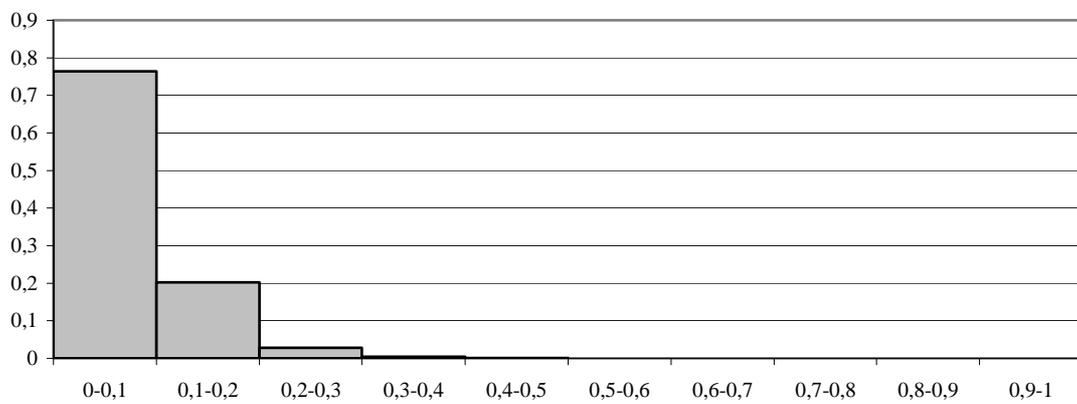
**IPG (errori di tipo Uniforme, NPC ranking)**



**Is (errori di tipo Uniforme, media)**



**Is (errori di tipo Uniforme, NPC ranking)**



I grafici sopra riportati rappresentano la suddivisione in percentuale delle graduatorie finali ottenute con la media e con l’NPC ranking (**asse delle ordinate**), in base ai valori dell’ $I_{PG}$  e dell’indice  $I_S$  calcolati su di esse (**asse delle ascisse**). In tutti i grafici si può notare come la maggior parte degli indici  $I_S$  e  $I_{PG}$  si distribuisca nell’intervallo [0-0.1]. Questo significa che gli errori generati dalla distribuzione Uniforme, analogamente alla distribuzione Normale, hanno modificato di poco i giudizi simulati su cui si è calcolato l’algoritmo e, di conseguenza, le classifiche finali ottenute sono molto simili a quella di riferimento. In questo contesto l’utilizzo della media risulta una buona soluzione per l’aggregazione dei giudizi di preferenza.

#### 5.4.4 Comportamento con errori di tipo Binomiale

Una variabile casuale discreta  $\varepsilon$  che assume valori interi positivi da 0, 1, ..., n si definisce Binomiale di indice n e parametro p con  $0 < p < 1$  se:

$$\Pr\{\varepsilon = t\} = \binom{n}{t} p^t (1-p)^{n-t} \quad \text{per } t = 0, 1, \dots, n.$$

Il fatto che  $\varepsilon$  abbia tale distribuzione è indicato in breve con la notazione  $\varepsilon \sim \text{Bi}(n, p)$ .

In particolare nelle simulazioni in SAS, per la generazione di dati, si sono usati come parametri  $n=8$  (n° di cartellini) e  $p=0.5$ :

```
ncas[i, j]=rel[i]+ranbin(&r, &k, 0.5)
```

dove “rel[i]” è il vettore della graduatoria di riferimento ed “&r” è la componente casuale “seed” e &k=8.

I risultati ottenuti dall’analisi sono riassunti nella seguente tabella:

$$1) \frac{\#(T_p \leq T_m)}{1000} = 0.559$$

$$\frac{\#(T_m \leq T_p)}{1000} = 0.487$$

$$2) \frac{\#(S_p \leq S_m)}{1000} = 0.6$$

$$\frac{\#(S_m \leq S_p)}{1000} = 0.436$$

$$3) \frac{\#(T_p \leq S_p)}{1000} = 0.844$$

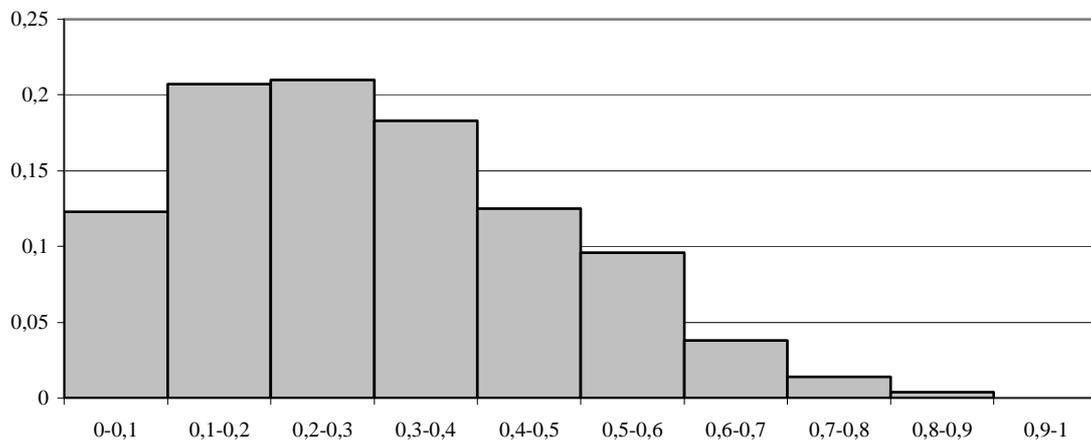
$$\frac{\#(S_p \leq T_p)}{1000} = 0.196$$

$$4) \frac{\#(T_m \leq S_m)}{1000} = 0.906$$

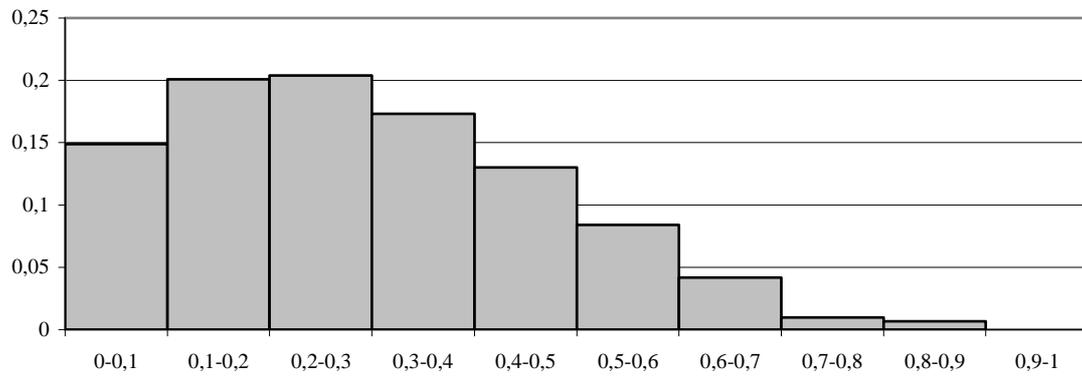
$$\frac{\#(S_m \leq T_m)}{1000} = 0.112$$

Osservando gli indici, si può notare che nel confronto (1) l'NPC ranking è leggermente migliore della media; questa tesi è rafforzata anche dai risultati ottenuti dal confronto (2) con l'indice di Spearman  $I_S$ . Mentre dall'analisi dei due indici di confronto emerge che l'indicatore  $I_{PG}$  è molto più potente dell'indice  $I_S$ .

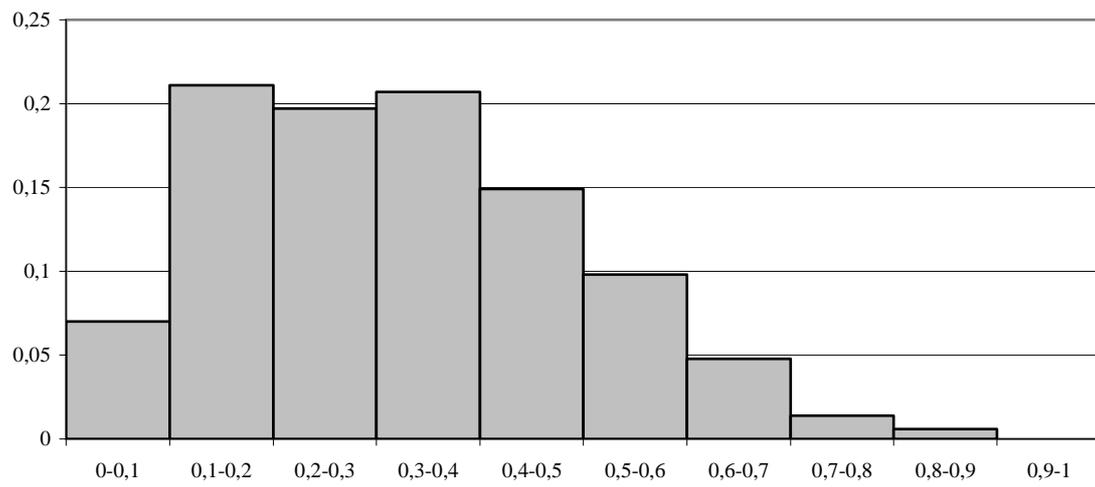
#### **$I_{PG}$ (errori di tipo Binomiale, media)**



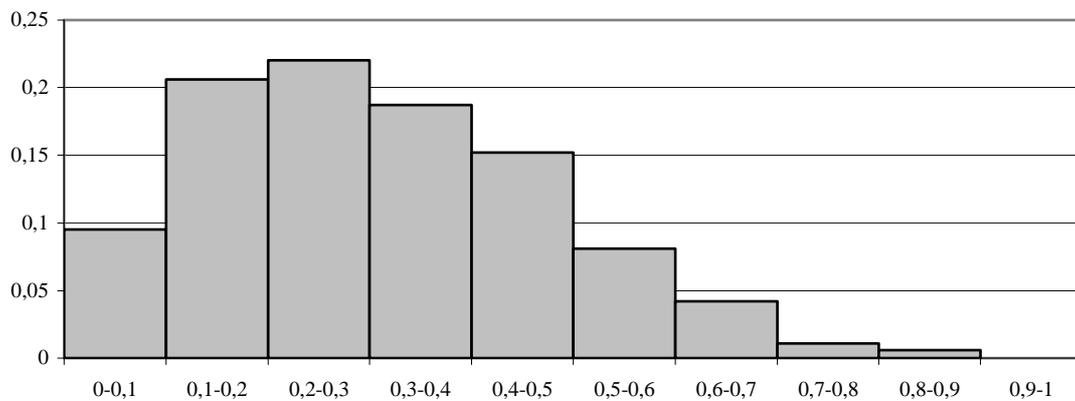
**IPG (errori di tipo Binomiale, NPC ranking)**



**Is (errori di tipo Binomiale, media)**



### Is (errori di tipo Binomiale, NPC ranking)



I grafici sopra riportati rappresentano la suddivisione in percentuale delle graduatorie finali ottenute con la media e con l’NPC ranking (**asse delle ordinate**), in base ai valori dell’indicatore  $I_{PG}$  e dell’indice  $I_S$  calcolati su di esse (**asse delle ascisse**). Nei due grafici in cui le classifiche sono state ottenute con l’NPC ranking si può notare che l’80% degli indici  $I_S$  e  $I_{PG}$  cade negli intervalli compresi tra 0 e 0.5 con almeno il 10% su ognuno di essi. Mentre nei due grafici in cui le classifiche sono state ottenute con la media gli intervalli più significativi sono compresi tra 0.1 e 0.6 nel caso dell’ $I_{PG}$  e tra 0 e 0.6 nel caso dell’ $I_S$ . Questo significa che gli errori generati dalla distribuzione Binomiale hanno modificato di molto i giudizi simulati su cui si è calcolato l’algoritmo, e, di conseguenza, le classifiche finali ottenute risultino distanti da quella di riferimento. In questo contesto l’utilizzo dell’NPC ranking per l’aggregazione dei giudizi di preferenza risulta migliore della media.

### 5.5 Studio di simulazione per i test TPG e TS

Prendendo sempre la stessa graduatoria di riferimento (1 2 3 4 5 6 7 8), si sono calcolate tutte le possibili permutazioni, date da  $8! = 40320$ . Successivamente, si sono analizzate per studiare quanto il test di permutazione TPG e un particolare test di Spearman TS siano sensibili ai cambiamenti delle classifiche finali osservate rispetto a quella di riferimento.

Cart1	Cart2	Cart3	Cart4	Cart5	Cart6	Cart7	Cart8	I <sub>S</sub>	I <sub>PG</sub>
1	2	3	4	5	6	7	8	0,00000000	0,00000000
1	2	3	4	5	6	8	7	0,01190476	0,00714286
1	2	3	4	5	7	6	8	0,01190476	0,00714286
1	2	3	4	5	7	8	6	0,03571429	0,02857143
1	2	3	4	5	8	6	7	0,03571429	0,02857143
1	2	3	4	5	8	7	6	0,04761905	0,03571429
1	2	3	4	6	5	7	8	0,01190476	0,00714286
1	2	3	4	6	5	8	7	0,02380952	0,01428571
1	2	3	4	6	7	5	8	0,03571429	0,02857143
1	2	3	4	6	7	8	5	0,07142857	0,06428571
1	2	3	4	6	8	5	7	0,05952381	0,05000000
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
4	8	7	6	5	2	3	1	0,86904762	0,87857143
4	8	7	6	5	3	1	2	0,86904762	0,87857143
4	8	7	6	5	3	2	1	0,88095238	0,88571429
5	1	2	3	4	6	7	8	0,11904762	0,11428571
5	1	2	3	4	6	8	7	0,13095238	0,12142857
5	1	2	3	4	7	6	8	0,13095238	0,12142857
5	1	2	3	4	7	8	6	0,15476191	0,14285714
5	1	2	3	4	8	6	7	0,15476191	0,14285714
5	1	2	3	4	8	7	6	0,16666667	0,15000000
5	1	2	3	6	4	7	8	0,14285714	0,13571429
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
8	7	6	5	3	2	1	4	0,92857143	0,93571429
8	7	6	5	3	2	4	1	0,96428571	0,97142857
8	7	6	5	3	4	1	2	0,97619048	0,98571429
8	7	6	5	3	4	2	1	0,98809524	0,99285714
8	7	6	5	4	1	2	3	0,95238095	0,96428571
8	7	6	5	4	1	3	2	0,96428571	0,97142857
8	7	6	5	4	2	1	3	0,96428571	0,97142857
8	7	6	5	4	2	3	1	0,98809524	0,99285714
8	7	6	5	4	3	1	2	0,98809524	0,99285714
8	7	6	5	4	3	2	1	1,00000000	1,00000000

Si è calcolato il test **TPG** per l'indicatore I<sub>PG</sub> nel seguente modo:

$$\mathbf{TPG} = \frac{\#(T^{oss} \geq T^*)}{40320}$$

dove  $T^{oss}$  si riferisce al valore assunto dall' $I_{PG}$  calcolato su una graduatoria scelta tra tutte quelle possibili, e  $T^*$  a tutti i valori assunti dall' $I_{PG}$  calcolato su tutte le possibili permutazioni.

Nello stesso modo si è calcolato il test **TS** per l'indice  $I_S$ :

$$TS = \frac{\#(T^{oss} \geq T^*)}{40320}$$

dove  $T^{oss}$  si riferisce al valore assunto dall'indice  $I_S$  calcolato su una graduatoria scelta tra tutte quelle possibili, e  $T^*$  a tutti i valori assunti dall' $I_S$  calcolato su tutte le possibili permutazioni.

Riportiamo l'analisi effettuata su alcune permutazioni:

Cart1	Cart2	Cart3	Cart4	Cart5	Cart6	Cart7	Cart8	$I_S$	$I_{PG}$	TS	TPG
1	2	3	4	5	6	7	8	0,000000	0,000000	0,000024	0,000024
1	2	3	4	5	6	8	7	0,011905	0,007143	0,000198	0,000198
1	2	3	4	8	7	6	5	0,119048	0,100000	0,018378	0,017535
3	2	1	6	7	8	4	5	0,226190	0,207143	0,075575	0,079613
4	3	7	2	6	1	8	5	0,392857	0,378571	0,291070	0,279510
2	8	5	6	1	4	3	7	0,488095	0,485714	0,467440	0,469970
4	3	7	5	1	8	6	2	0,500000	0,500000	0,511610	0,508280
5	8	1	3	7	6	4	2	0,630952	0,635714	0,731800	0,731500
6	5	2	8	4	7	3	1	0,702381	0,707143	0,850400	0,841370
7	4	8	6	5	1	3	2	0,869048	0,885714	0,981620	0,978450
8	7	6	5	2	3	1	4	0,916667	0,928571	0,994640	0,991870
8	7	6	5	4	3	2	1	1,000000	1,000000	1,000000	1,000000

Osservando la tabella sopra riportata si può notare come TS e TPG unitamente agli indici  $I_S$  e  $I_{PG}$  si allontanino da zero all'aumentare della distanza della classifica osservata rispetto a quella di riferimento.

## 5.6 Potenza dei test sotto $H_1$

Dopo aver analizzato i due test sotto  $H_0$  (avente come classifica di riferimento 1 2 3 4 5 6 7 8), si è presa in considerazione come ipotesi alternativa  $H_1$  in cui la nuova graduatoria di riferimento è: 1 2 3 6 4 5 7 8. Si sono ripetute le mille simulazioni precedentemente descritte generando degli errori casuali di tipo: Normale, Gamma,

Uniforme e Binomiale. Successivamente si sono calcolati per ogni distribuzione mille test di Spearman per le medie ( $TS_M$ ), mille per l’NPC ranking ( $TS_{NPC}$ ), mille test di permutazione per le medie ( $TPG_M$ ) e mille per l’NPC ranking ( $TPG_{NPC}$ ). Per saggiare la potenza dei due test si sono calcolate le percentuali cumulate di classifiche finali aventi un determinato valore dei due test, ottenendo i seguenti risultati:

Distr:	NORMALE				UNIFORME			
$\alpha$	$TS_M$	$TS_{NPC}$	$TPG_M$	$TPG_{NPC}$	$TS_M$	$TS_{NPC}$	$TPG_M$	$TPG_{NPC}$
<b>0.010</b>	0,030	0,032	0,032	0,038	0,163	0,135	0,163	0,135
<b>0.025</b>	0,196	0,212	0,206	0,220	0,464	0,384	0,493	0,401
<b>0.050</b>	0,716	0,634	0,728	0,672	0,776	0,656	0,795	0,672
<b>0.075</b>	0,896	0,838	0,896	0,836	0,907	0,811	0,907	0,813
<b>0.100</b>	0,956	0,886	0,958	0,888	0,943	0,870	0,948	0,878
<b>0.200</b>	1,000	0,996	1,000	0,996	0,996	0,981	0,997	0,986
<b>0.300</b>	1,000	1,000	1,000	1,000	0,999	0,994	0,999	0,995
<b>0.400</b>	1,000	1,000	1,000	1,000	1,000	0,998	1,000	0,998
<b>0.500</b>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
<b>0.600</b>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
<b>0.700</b>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
<b>0.800</b>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
<b>0.900</b>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
<b>1.000</b>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Distr:	BINOMIALE				GAMMA			
Intervalli	$TS_M$	$TS_{NPC}$	$TPG_M$	$TPG_{NPC}$	$TS_M$	$TS_{NPC}$	$TPG_M$	$TPG_{NPC}$
<b>0.010</b>	0,492	0,408	0,506	0,409	0,017	0,020	0,017	0,020
<b>0.025</b>	0,924	0,860	0,933	0,869	0,051	0,053	0,062	0,059
<b>0.050</b>	0,994	0,985	0,998	0,987	0,137	0,150	0,147	0,160
<b>0.075</b>	1,000	0,998	1,000	0,999	0,218	0,220	0,220	0,221
<b>0.100</b>	1,000	0,999	1,000	0,999	0,286	0,272	0,304	0,279
<b>0.200</b>	1,000	1,000	1,000	1,000	0,525	0,480	0,558	0,504
<b>0.300</b>	1,000	1,000	1,000	1,000	0,652	0,625	0,675	0,647
<b>0.400</b>	1,000	1,000	1,000	1,000	0,772	0,751	0,774	0,755
<b>0.500</b>	1,000	1,000	1,000	1,000	0,841	0,830	0,842	0,831
<b>0.600</b>	1,000	1,000	1,000	1,000	0,898	0,883	0,902	0,895
<b>0.700</b>	1,000	1,000	1,000	1,000	0,936	0,928	0,937	0,928
<b>0.800</b>	1,000	1,000	1,000	1,000	0,963	0,962	0,964	0,964
<b>0.900</b>	1,000	1,000	1,000	1,000	0,986	0,984	0,987	0,984
<b>1.000</b>	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Osservando la tabella si può notare che i due test TS e TPG risultano molto potenti in quanto segnalano già ad  $\alpha$  nominali bassi che la classifica utilizzata per le simulazioni sotto l’ipotesi alternativa è diversa da quella di riferimento.



## **Capitolo 6**

### ***APPLICAZIONI ALLO STUDIO PILOTA DI CONJOINT ANALYSIS SUL PRODOTTO BANCARIO: CONTO GIOVANI***

#### **6.1 Indagine preliminare**

In questo capitolo si presenta uno studio pilota condotto su un particolare prodotto bancario: il conto giovani, intervistando un campione di giovani piuttosto contenuto (poco più di 50 persone), compreso tra i 18 e i 29 anni.

L'obiettivo è di comprendere se il settore bancario italiano può essere realmente interessato a sviluppare eventuali progetti sfruttando le potenzialità della Conjoint Analysis, tenendo conto che in letteratura sono presenti già numerosi casi, soprattutto condotti negli Stati Uniti, in cui si nota che la C.A. può portare a risultati convincenti.

#### **6.2 Svolgimento dell'indagine**

Il primo passo per lo svolgimento dell'indagine è stato quello di raccogliere quante più informazioni possibili sul servizio "conto giovani", attualmente offerto dai principali istituti bancari veneti, e tutti i vari aspetti come ad es. agevolazioni, costi, o servizi extra, offerti alla clientela da questo particolare conto.

Una volta raccolte queste informazioni si è cercato di individuare le caratteristiche principali di questo servizio (definiti nel gergo dalla Conjoint Analysis "attributi") e le varie opzioni con cui esse sono disponibili (ovvero i "livelli"). Sono stati individuati

cinque attributi su due livelli. Di seguito si riporta la tabella con gli attributi e i relativi livelli trovati:

<b><u>ATTRIBUTI</u></b>	<b>COSTO</b>	<b>AGEVOL. FINANZ.</b>	<b>CARTE</b>	<b>POLIZZE</b>	<b>SERVIZI EXTRA</b>
<b><u>LIVELLI</u></b>	<ul style="list-style-type: none"> <li>• Fisso al mese da 6 a 10 euro con operazioni gratuite illimitate</li> <li>• Fisso al mese da 0 a 5 euro con operazioni gratuite limitate</li> </ul>	<ul style="list-style-type: none"> <li>• Mutuo e/o prestiti agevolati</li> <li>• Agevolazioni su investim.</li> </ul>	<ul style="list-style-type: none"> <li>• Bancomat gratuito e bancomat internaz. con costo fisso annuo</li> <li>• Bancomat gratuito e carta di credito con costo fisso annuo</li> </ul>	<ul style="list-style-type: none"> <li>• Furto e/o infortuni</li> <li>• Responsabilità civile</li> </ul>	<ul style="list-style-type: none"> <li>• Sconti con negozi, abbonamenti, riviste</li> <li>• Convenzioni viaggi, Telepass</li> </ul>

**tabella 6.1**

In comune accordo con le banche si è deciso di presentare i seguenti otto cartellini.

➤ **Cartellino 1**

Costo Fisso minore + operazioni limitate 0-5

Agevolazioni finanziarie Agevolazioni su pac/investimenti

Carte bancomat gratis+bancomat int. con fisso

Polizze Furto e/o infortuni

Servizi Extra Convenzioni viaggi, Telepass

➤ **Cartellino 2**

Costo Fisso+operazioni illimitate (6-10 euro)

Agevolazioni finanziarie Agevolazioni su pac/investimenti

Carte bancomat gratis+carta credito con fisso  
Polizza Furto e/o infortuni  
Servizi Extra Convenzioni viaggi, Telepass

➤ **Cartellino 3**

Costo Fisso+operazioni illimitate (6-10 euro)  
Agevolazioni finanziarie Mutuo e/o prestiti agevolati  
Carte bancomat gratis+bancomat int. con fisso  
Polizze Furto e/o infortuni  
Servizi Extra Sconti con negozi /abbonamento riviste

➤ **Cartellino 4**

Costo Fisso+operazioni illimitate (6-10 euro)  
Agevolazioni finanziarie Mutuo e/o prestiti agevolati  
Carte bancomat gratis+carta credito con fisso  
Polizze Responsabilità civile  
Servizi Extra Convenzioni viaggi, Telepass

➤ **Cartellino 5**

Costo Fisso+operazioni illimitate (6-10 euro)  
Agevolazioni finanziarie Agevolazioni su pac/investimenti  
Carta bancomat gratis+bancomat int. con fisso  
Polizze Responsabilità civile  
Servizi Extra Sconti con negozi /abbonamento riviste

➤ **Cartellino 6**

Costo Fisso minore + operazioni limitate 0-5  
Agevolazioni finanziarie Mutuo e/o prestiti agevolati  
Carte bancomat gratis+carta credito con fisso  
Polizze Furto e/o infortuni  
Servizi Extra Sconti con negozi /abbonamento riviste

➤ **Cartellino 7**

Costo Fisso minore + operazioni limitate 0-5  
Agevolazioni finanziarie Agevolazioni su pac/investimenti  
Carte bancomat gratis+carta credito con fisso  
Polizze Responsabilità civile  
Servizi Extra Sconti con negozi /abbonamento riviste

➤ **Cartellino 8**

Costo Fisso minore + operazioni limitate 0-5  
Agevolazioni finanziarie Mutuo e/o prestiti agevolati  
Carte bancomat gratis+bancomat int. con fisso  
Polizze Responsabilità civile  
Servizi Extra Convenzioni viaggi, Telepass

Si è arricchito ogni cartellino da elementi grafici per rendere più gradevole l'intervista.  
Si riporta di seguito l'esempio di un cartellino:



**Conto Giovani**

- **Costo:** fisso al mese di 6 e 3 euro con operazioni gratuite illimitate
- **Agevolazioni finanziarie:** Agevolazioni su investimenti
- **Carte:** bancomat greenlife e bancomat international con costo fisso annuo
- **Polizze:** furto dell'identità
- **Servizi extra:** Convenzioni viaggi, Telepass

### **6.3 Modalità di intervista**

Nelle interviste si è chiesto al rispondente di dare un giudizio su una scala di punteggio tra 0 e 10 (dove 10 è meglio) a ciascun cartellino, assicurando in tal modo una maggiore omogeneità dei risultati. Si è concesso ai rispondenti anche la possibilità di dare punteggi con numeri decimali. A ciascun intervistato si è presentato prima un foglio con tutti i profili di servizio presenti, in modo che potesse farsi un'idea generale dell'argomento trattato, successivamente gli si sono presentati gli otto cartellini.

### **6.5 Scelta ed analisi del campione**

Il campione di popolazione che si è intervistato, è costituito da maggiorenni fino a 29 anni. Si sono individuati cinque diverse tipologie di potenziali clienti del conto giovani:

- Studenti (12 intervistati)
- Studenti lavoratori (10 intervistati)
- Lavoratori dipendenti (13 intervistati)
- Lavoratori autonomi (10 intervistati)
- Disoccupati (11 intervistati)

TOTALE: 56 interviste

Prima di svolgere l'intervista vera e propria è stato chiesto a ciascun rispondente di fornire i propri dati anagrafici, in tal modo è possibile analizzare i risultati non solo del campione intero ma anche in base alla variabile che si desidera (sesso, età, etc..). In questa tesi si è analizzato solo il campione totale.

In dettaglio è stato richiesto:

- Et : .....
- Sesso:        M<sup>1</sup>    F<sup>1</sup>
- Residenza: .....
- Titolo di studio: .....
- Attivit :    1 Studente universitario  
              1 Studente universitario e lavoratore  
              1 Lavoratore dipendente  
              1 Lavoratore autonomo  
              1 In cerca di occupazione

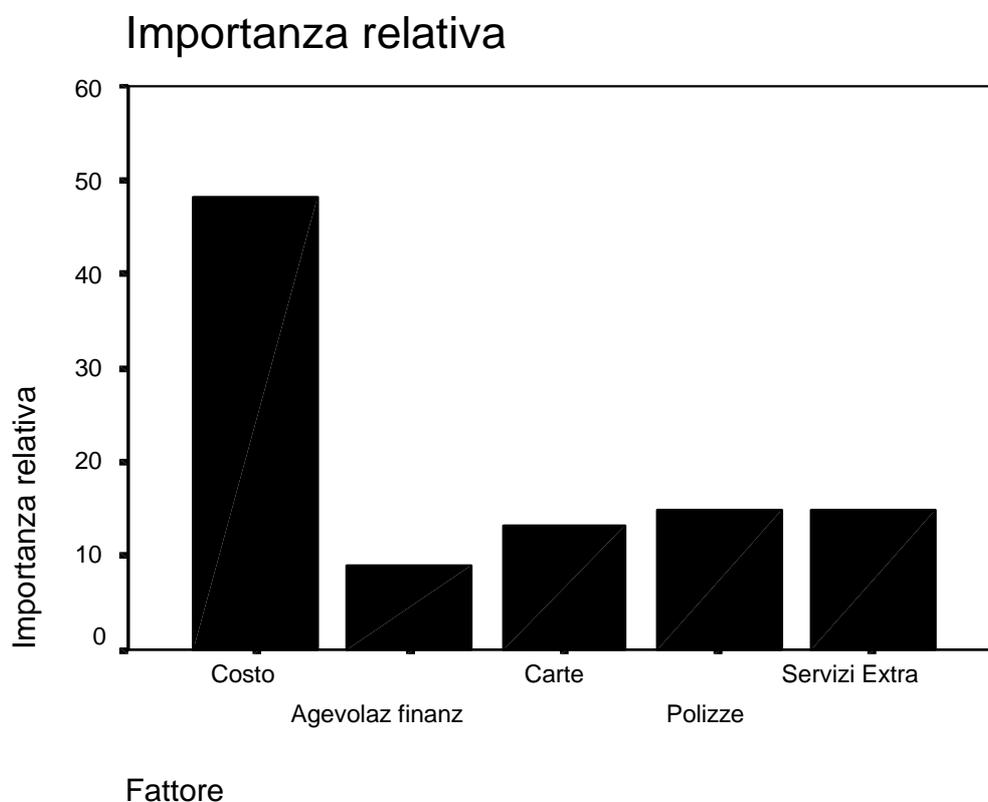
Si riportano di seguito i risultati delle interviste, ovvero i punteggi assegnati dai rispondenti agli otto cartellini proposti (c1,c2,...,c8). A titolo esemplificativo sono indicati soltanto i punteggi assegnati dai primi cinque rispondenti per ciascun segmento di clientela.

gruppo	sezzo	età	titolo di studio	c1	c2	c3	c4	c5	c6	c7	c8
studente	F	25	media sup.	5	10	6,5	7,5	8,5	4	1	0
	M	21	media sup.	4	9	6	10	8	0	2	3
	F	22	media sup.	2	10	5,5	8	7,5	1	0	3,5
	M	24	dip univ.	0	10	5	8,5	7	1,5	3	4
	M	25	media sup.	3	10	6,5	9	7	5	0	2
studente lavoratore	F	22	media sup.	4	10	6,5	7	9	5	1	0
	F	24	media sup.	3,5	10	6	8,5	7,5	5	0	2
	M	24	media sup.	3	10	5	8	6	7	0	4,5
	F	27	media sup.	2	9	4,5	10	7	5	0	4
	M	28	media sup.	4	10	5	8	7	6	0	3
lavoratore dipendente	F	23	media sup.	0	6	8,5	10	7	4	2	5
	M	26	laurea	5,5	10	8	7	6	4	0	3
	F	25	media sup.	4	7	10	9	7,5	6,5	0	6
	F	24	dipl. univ.	8	10	0	4	5	7	6,5	6
	M	25	media sup.	5	6	10	8,7	7	4	0	3
lavoratore autonomo	M	28	media sup.	10	8	3	2	0	1	5	7
	M	29	laurea	10	7	3	1	0	2	5	9
	M	29	media sup.	7	3	2	9	0	6	5	10
	M	29	laurea	2	10	6	7	5	3	0	1
	M	29	dipl. univ.	9	4	0	6	2	7	8	10
disoccupato	M	26	laurea	6,5	3	5	5,5	0	8	6	10
	F	26	media sup.	8	2	6,5	3	0	10	4,5	5,5
	M	24	media sup.	6	2	4	5	0	8	3	10
	F	29	laurea	8,5	2	4,5	3	0	10	7	9,5
	F	23	media sup.	5	2	4	6	0	9	7	10

Concluse le interviste, si sono analizzati i giudizi degli intervistati in SAS con l'NPCC ranking per ottenere la classifica finale dei vari cartellini e poi si è proseguita l'analisi delle singole utilità con il software SPSS Conjoint. L'NPC ranking ha stabilito che il secondo cartellino è quello preferito dagli intervistati, mentre quello peggiore è il sesto. Per quanto riguarda l'SPSS, per ottenere una procedura di Conjoint Analysis, è necessario specificare determinati comandi, ovvero utilizzare una sintassi particolare in modo da ottenere un'analisi su misura, in base alle esigenze dello studio in questione. L'analisi in SPSS richiede la definizione di un "plan file", creato usando la procedura per la generazione del piano ortogonale. Il "plan file" è il file che raccoglie i valori numerici assegnati ai livelli dei vari attributi. In questo caso, non essendoci livelli quantificabili numericamente, si sono assegnati i valori di +1 e -1 per le coppie di livelli di ciascun attributo. Il "plan file" quindi è il seguente:

A	B	C	D	E	CARD_
1,00	1,00	-1,00	-1,00	1,00	1
-1,00	1,00	1,00	-1,00	1,00	2
-1,00	-1,00	-1,00	-1,00	-1,00	3
-1,00	-1,00	1,00	1,00	1,00	4
-1,00	1,00	-1,00	1,00	-1,00	5
1,00	-1,00	1,00	-1,00	-1,00	6
1,00	1,00	1,00	1,00	-1,00	7
1,00	-1,00	-1,00	1,00	1,00	8

Grazie ai risultati ottenuti, si può capire per ogni attributo, qual è il livello preferito dai rispondenti e qual è l'importanza relativa assegnata implicitamente dagli intervistati a ciascun attributo. Di seguito viene riportato il grafico che riguarda l'utilità l'importanza relativa degli attributi.



Studiando il grafico dell'importanza relativa, si nota come l'attributo "costo" viene considerato quello più importante. Seguono, per ordine di importanza, con la stessa percentuale, gli attributi "servizi extra" e le "polizze", e infine le "carte" e le "agevolazioni finanziarie".

In conclusione le banche, coinvolte nello studio pilota sul "conto giovani", si sono dimostrate soddisfatte dei risultati ottenuti ed interessate ad eventuali nuovi progetti di indagine sfruttando le potenzialità della C.A..



## APPENDICE

### SAS MACRO LANGUAGE

#### A) Metodologia NPC ranking

```
libname pesi 'percorso';
data pesi;
infile 'C :percorso ';
input w1;
run;
libname mediag 'C: percorso;
data mediag;
infile 'C: percorso ;
input m1;
run;
%macro graduatoria(w,u,n,k);
/* generazione di dati */
proc iml;
use &u;
read all var _num_ into x;
re1=x;
ncas=j(&k,&n,1);
do i=1 to &k;
do j=1 to &n;
ncas[i,j]=sqrt(3)*normal(5733345)+re1[i];
end;
end;
create ncasuale from ncas;
append from ncas;
quit;
run;
/* ordina la matrice secondo la procedura rank */
proc rank data=ncasuale ties=high out=rank;
run;
/* procedura per la matrice R delle classifiche parziali */
data prv;
set rank;
array prov(&n) col1-col&n;
array new(&n) a1-a&n;
do i=1 to &n;
new(i)=(prov(i)+0.5)/(&k+1);
end;
data R;
set prv (keep=a1-a&n);
run;
/* Funzione di Fisher */
data fisher;
set R;
```

```

array prov(&n) a1-a&n;
array new(&n) b1-b&n;
do i=1 to &n;
new(i) =-log(1-prov(i));
end;
run;
data Z;
set fisher (keep=b1-b&n);
run;
proc iml;
use Z;
read all var _num_ into x;
use &w;
read all var _num_ into y;
h=x*y;
create re from h;
append from h;
quit;
run;
proc rank data=re ties=high out=rank1;
run;
/* si ottiene la classifica finale per l’NPC ranking */
data clasfin;
set rank1;
graduatoria=(COL1)/(&k);
data finale;
set clasfin(keep = graduatoria);
run;
% mend;
% graduatoria(w=pesi,u=mediag,n=5,k=4);

```

## B) Media aritmetica

```

proc iml;
use ncasuale;
read all var _num_ into x;
use &w;
read all var _num_ into y;
m=(x*y);
create med from m;
append from m;
quit;
run;
data media;
set med;
classifica=(COL1)/(&n);
data medie;
set media(keep = classifica);
proc rank data=medie ties=high out=rankm;
run;

```

### C) Indicatore $I_{PG}$

```
/* matrice km*/
proc iml;
use "graduatoria ottenuta con l'NPC ranking o con la media";
read all var _num_ into x;
matm=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
if x[i]>x[j]
then matm[i,j-1]=1;
end;
end;
create Km from matm;
append from matm;
quit;
run;
/* matrice Lm*/
proc iml;
use &u;
read all var _num_ into x;
matlm=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
matlm[i,j-1]=abs(x[i]-x[j])-1;
end;
end;
create Lm from matlm;
append from matlm;
quit;
run;
/* matrice H*/
proc iml;
use rankm;
read all var _num_ into x;
matH=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
math[i,j-1]=abs(x[i]-x[j])-1;
end;
end;
create H from math;
append from math;
quit;
run;
/* calcolo della statistica test */
proc iml;
use Km;
read all var _num_ into x;
use Lm;
read all var _num_ into y;
use H;
read all var _num_ into z;
mat1=j(&k-1,&k-1,0);
mat2=j(&k-1,&k-1,0);
v2=j(&k-1,1,1);
v1=j(1,&k-1,1);
do i=1 to &k-1;
```

```

do j=1 to &k-1;
mat1[i,j]=x[i,j]*y[i,j];
mat2[i,j]=x[i,j]*z[i,j];
end;
end;
T2=v1*x*v2+v1*mat1*v2+v1*mat2*v2;
create Tm2 from T2;
append from T2;
quit;
run;
data Tm1;
set Tm2;
Tm=col1;
run;
data Tm;
set Tm1(keep=Tm);
run;
/* calcolo dell'indicatore Ipg*/
proc iml;
use Tm;
read all var _num_ into x;
den=(&k*(&k*&k-1))/3;
deno=den-(&k*(&k-1)/2);
max1=x/deno;
create maxm from max1;
append from max1;
quit;
run;

```

## D) Indice di Spearman I<sub>s</sub>

```

proc iml;
use &u;
read all var _num_ into x;
use "graduatoria ottenuta con l'NPC ranking o con la media";
read all var _num_ into y;
s=(y-x);
create s1 from s;
append from s;
quit;
run;
proc iml;
use s1;
read all var _num_ into x;
dif=x;
quadr=j(&k,1,0);
do i=1 to &k;
quadr[i,1]=dif[i]*dif[i];
end;
create dquad from quadr;
append from quadr;
quit;
run;
proc iml;
use dquad;
read all var _num_ into q1;
sommam=3*sum(q1);
create s1a from sommam;

```

```

append from sommam;
quit;
run;
data sp1;
set s1a;
Spm=(COL1/(&k*(&k*&k-1)));
data spearmanm;
set sp1(keep =Spm);
run;

```

## E) $I_{PG}$ e $I_S$ calcolati su tutte le Permutazioni

```

libname mediag 'C: ';
data mediag;
infile 'C: ';
input m1;
run;
%let e=k!;
%macro permutazioni(u,k);
proc iml;
g=j(1,1,&k);
do i=1 to &k-1;
f=(&k-i);
g=g*f;
end;
create kf from g;
append from g;
quit;
run;
proc plan seed= "seed ";
factors subject=&e
order=&k ordered;
treatments stimulus=&k perm;
output out=psy;
run;
data permut;
set psy(keep=stimulus);
run;
proc iml;
use permut(firstobs=1 obs=&e);
read all var _num_ into x;
use kf;
read all var _num_ into y;
mat=j(y,&k,1);
do i=1 to y*&k;
mat[i]=x[i];
end;
create permu from mat;
append from mat;
quit;
run;
proc transpose data=permu out=permutt;
run;
proc iml;
use permutt;
read all var _num_ into x;
use kf;

```

```

read all var _num_ into y;
a=j(&k-1,&k-1,0);
do z=1 to y;
mat=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
if x[i,z]>x[j,z]
then mat[i,j-1]=1;
end;
end;
a=a//mat;
end;
create a1 from a;
append from a;
quit;
run;
proc iml;
use permutt;
read all var _num_ into x;
use kf;
read all var _num_ into y;
b=j(&k-1,&k-1,0);
do z=1 to y;
mat1=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
mat1[i,j-1]=abs(x[i,z]-x[j,z])-1;
end;
end;
b=b//mat1;
end;
create b1 from b;
append from b;
quit;
run;
proc iml;
use &u;
read all var _num_ into x;
math=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
math[i,j-1]=abs(x[i]-x[j])-1;
end;
end;
create H1 from math;
append from math;
quit;
run;
proc iml;
use H1;
read all var _num_ into x;
use kf;
read all var _num_ into y;
H2=j(&k-1,&k-1,0);
do i=1 to y;
H2=H2//x;
end;
create H from H2;

```

```

append from H2;
quit;
run;
proc iml;
use &u;
read all var _num_ into x;
use permutt;
read all var _num_ into y;
use kf;
read all var _num_ into z;
mat=j(&k,1,0);
do i=1 to z;
s=(y[,i]-x);
mat=mat||s;
end;
create s1 from mat;
append from mat;
quit;
run;
proc iml;
use s1;
read all var _num_ into x;
dif=x;
use kf;
read all var _num_ into z;
quadr=j(&k,1,0);
dat=j(&k,1,0);
do j=2 to z+1;
do i=1 to &k;
quadr[i,1]=dif[i,j]*dif[i,j];
end;
dat=dat||quadr;
end;
create dquad from dat;
append from dat;
quit;
run;
proc iml;
use dquad;
read all var _num_ into q1;
use kf;
read all var _num_ into z;
som=j(1,1,0);
do i=2 to z+1;
sommam=3*sum(q1[,i]);
som=som||sommam;
end;
create s1a from som;
append from som;
quit;
run;
proc iml;
use s1a;
read all var _num_ into x;
use kf;
read all var _num_ into z;
sper=j(1,1,0);
do i=2 to z+1;

```

```

t=(x[,i]/(&k*(&k*&k-1)));
sper=sper//t;
end;
create spear from sper;
append from sper;
quit;
run;
proc iml;
use spear;
read all var _num_ into sp;
use max;
read all var _num_ into maxT;
conf=sp||maxT;
create confront1 from conf;
append from conf;
quit;
data confront3;
set confront1(firstobs=2 obs=&e+1);
run;
data confront2;
set confront3;
spearman=col1;
maxT=col2;
run;
data confronto1;
set confront2(keep=spearman maxT);
proc print data=confronto1;
run;
%mend;
%permutazioni(u=mediag,k=8)

```

#### *F) Studio di simulazione*

```

libname pesi 'C:\percorso';
data pesi;
infile 'C:\percorso\pesi1.txt';
input w1;
run;
libname mediag 'C:\percorso';
data mediag;
infile 'C:\percorso\medie.txt';
input m1;
run;
libname prova 'C:\percorso';
data prova;
infile 'C:\percorso\moreno.txt';
input pm pp sm sp;
run;
data zero;
infile 'C:\percorso\zero.txt';
input z1;
run;
%macro graduatoria(w,u,mr,ze,n,k);
%do o=1 %to 1000;
/* generazione di dati */
%let r=%eval(124+20*&o+1);
proc iml;
use &ze;
read all var _num_ into x;

```

```

re1=x;
ncas=j(&k,&n,1);
do i=1 to &k;
do j=1 to &n;
    • ncas[i,j]=sqrt(1)*normal(&r)+re1[i]; Distr. Normale
    • ncas[i,j]= x[i]+rangam(&r,1); Distr. Gamma
    • ncas[i,j]=re1[i]+ranuni(&r); Distr. Uniforme
    • ncas[i,j]=re1[i]+ranbin(&r,&k,0.5); Distr. Binomiale

end;
end;
create ncasuale from ncas;
append from ncas;
quit;
run;
/* ordina la matrice secondo la procedura rank */
proc rank data=ncasuale ties=high out=rank;
proc print data=rank;
run;
/* procedura per la matrice R delle classifiche parziali */
data prv;
set rank;
array prov(&n) col1-col&n;
array new(&n) a1-a&n;
do i=1 to &n;
new(i)=(prov(i)+0.5)/(&k+1);
end;
data R;
set prv (keep=a1-a&n);
run;
/* Funzione di Fisher */
data fisher;
set R;
array prov(&n) a1-a&n;
array new(&n) b1-b&n;
do i=1 to &n;
new(i)=-log(1-prov(i));
end;
run;
data Z;
set fisher (keep=b1-b&n);
run;
proc iml;
use Z;
read all var _num_ into x;
use &w;
read all var _num_ into y;
h=x*y;
create re from h;
append from h;
quit;
run;
/* ordina il vettore degli Zi secondo la procedura rank */
proc rank data=re ties=high out=rank1;
run;
/* si ottiene la classifica finale */
data clasfin;
set rank1;

```

```

graduatoria=(COL1)/(&k);
data finale;
set clasfin(keep = graduatoria);
run;
/* procedura per le medie */
proc iml;
use ncasuale;
read all var _num_ into x;
use &w;
read all var _num_ into y;
m=(x*y);
create med from m;
append from m;
quit;
run;
data media;
set med;
classifica=(COL1)/(&n);
data medie;
set media(keep = classifica);
proc rank data=medie ties=high out=rankm;
run;
/* IPG per la media */
/* matrice km*/
proc iml;
use rankm;
read all var _num_ into x;
matm=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
if x[i]>x[j]
then matm[i,j-1]=1;
end;
end;
create Km from matm;
append from matm;
quit;
run;
/* matrice Lm*/
proc iml;
use rankm;
read all var _num_ into x;
matlm=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
matlm[i,j-1]=abs(x[i]-x[j])-1;
end;
end;
create Lm from matlm;
append from matlm;
quit;
run;
/* matrice H*/
proc iml;
use &u;
read all var _num_ into x;
matH=j(&k-1,&k-1,0);
do i=1 to &k-1;

```

```

do j=i+1 to &k;
math[i,j-1]=abs(x[i]-x[j])-1;
end;
end;
create H from math;
append from math;
quit;
run;
proc iml;
use Km;
read all var _num_ into x;
use Lm;
read all var _num_ into y;
use H;
read all var _num_ into z;
mat1=j(&k-1,&k-1,0);
mat2=j(&k-1,&k-1,0);
v2=j(&k-1,1,1);
v1=j(1,&k-1,1);
do i=1 to &k-1;
do j=1 to &k-1;
mat1[i,j]=x[i,j]*y[i,j];
mat2[i,j]=x[i,j]*z[i,j];
end;
end;
T2=v1*x*v2+v1*mat1*v2+v1*mat2*v2;
create Tm2 from T2;
append from T2;
quit;
run;
data Tm1;
set Tm2;
Tm=col1;
run;
data Tm;
set Tm1(keep=Tm);
run;
proc iml;
use Tm;
read all var _num_ into x;
den=(&k*(&k*&k-1))/3;
deno=den-(&k*(&k-1)/2);
max1=x/deno;
create maxm from max1;
append from max1;
quit;
run;
/* IPG per l'NPC ranking */
/* matrice kp*/
proc iml;
use rank1;
read all var _num_ into x;
matp=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
if x[i]>x[j]
then matp[i,j-1]=1;
end;
end;

```

```

end;
create Kp from matp;
append from matp;
quit;
run;
/* matrice Lp*/
proc iml;
use rank1;
read all var _num_ into x;
matlp=j(&k-1,&k-1,0);
do i=1 to &k-1;
do j=i+1 to &k;
matlp[i,j-1]=abs(x[i]-x[j])-1;
end;
end;
create Lp from matlp;
append from matlp;
quit;
run;
proc iml;
use Kp;
read all var _num_ into x;
use Lp;
read all var _num_ into y;
use H;
read all var _num_ into z;
mat1=j(&k-1,&k-1,0);
mat2=j(&k-1,&k-1,0);
v2=j(&k-1,1,1);
v1=j(1,&k-1,1);
do i=1 to &k-1;
do j=1 to &k-1;
mat1[i,j]=x[i,j]*y[i,j];
mat2[i,j]=x[i,j]*z[i,j];
end;
end;
T2=v1*x*v2+v1*mat1*v2+v1*mat2*v2;
create Tp2 from T2;
append from T2;
quit;
run;
data Tp1;
set Tp2;
Tp=coll;
run;
data Tp;
set Tp1(keep=Tp);
run;
proc iml;
use Tp;
read all var _num_ into x;
den=(&k*(&k*&k-1))/3;
deno=den-(&k*(&k-1)/2);
max1=x/deno;
create maxp from max1;
append from max1;
quit;
run;

```

```

/*Is per la media*/
proc iml;
use &u;
read all var _num_ into x;
use rankm;
read all var _num_ into y;
s=(y-x);
create s1 from s;
append from s;
quit;
run;
proc iml;
use s1;
read all var _num_ into x;
dif=x;
quadr=j(&k,1,0);
do i=1 to &k;
quadr[i,1]=dif[i]*dif[i];
end;
create dquad from quadr;
append from quadr;
quit;
run;
proc iml;
use dquad;
read all var _num_ into q1;
sommam=3*sum(q1);
create s1a from sommam;
append from sommam;
quit;
run;
data sp1;
set s1a;
Spm=(COL1/(&k*(&k*&k-1)));
data spearmanm;
set sp1(keep =Spm);
run;
/*Is per l’NPC ranking*/
proc iml;
use &u;
read all var _num_ into x;
use rank1;
read all var _num_ into y;
s=(y-x);
create s2 from s;
append from s;
quit;
run;
proc iml;
use s2;
read all var _num_ into x;
dif2=x;
quadr2=j(&k,1,0);
do i=1 to &k;
quadr2[i,1]=dif2[i]*dif2[i];
end;
create dquad2 from quadr2;
append from quadr2;

```

```

quit;
run;
proc iml;
use dquad2;
read all var _num_ into q2;
somap=3*sum(q2);
create s2a from somap;
append from somap;
quit;
run;
data sp2;
set s2a;
Spp=(COL1/(&k*(&k*&k-1)));
data spearmanp;
set sp2(keep =Spp);
run;
proc iml;
use maxm;
read all var _num_ into pm;
use maxp;
read all var _num_ into pp;
use Spearmanm;
read all var _num_ into sm;
use Spearmanp;
read all var _num_ into sp;
e=pm||pp||sm||sp;
create e1 from e;
append from e;
quit;
run;
data fina;
set e1;
Pm=col1;
Pp=col2;
Sm=col3;
Sp=col4;
run;
data finale&o;
set fina(keep=pm pp sm sp);
run;
%end;
%do c=1 %to 1000;
proc datasets;
append base=mr data=finale&c;
%end;
proc print data=mr;
run;
%mend graduatoria;
%graduatoria(w=pesi,u=mediag,mr=prova,ze=zero,n=5,k=8);

```

## G) Confronto tra indici

```

proc iml;
use ;
read all var _num_ into x;
use ;

```

```

read all var _num_ into y;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=y[i]
then mt[i,1]=1;
end;
s=sum(mt);
r=s/1000;
print r;
quit;

```

## H) I test TS e TPG

```

libname perms 'C:\Percorso';
data perms;
infile 'C:\Percorso\perms.txt';
input ps;
run;
libname permm 'C:\Percorso';
data permm;
infile 'C:\Percorso\permm.txt';
input pm;
run;
libname Tm 'C:\Percorso';
data Tm;
infile 'C:\Percorso\Tm.txt';
input Tm;
run;
libname Tp 'C:\Percorso';
data Tp;
infile 'C:\Percorso\Tp.txt';
input Tp;
run;
data Sm;
infile 'C:\Percorso\sm.txt';
input Sm;
run;
data Sp;
infile 'C:\Percorso\sp.txt';
input Sp;
run;
proc iml;
use perms;
read all var _num_ into x;
vet=j(1008,40,0);
do i=1 to 40320;
vet[i]=x[i];
end;
create p from vet;
append from vet;
quit;
data milles;
set p(keep=col1);
run;
proc iml;
use permm;
read all var _num_ into y;
vet1=j(1008,40,0);

```

```

do i=1 to 40320;
vet1[i]=y[i];
end;
create p1 from vet1;
append from vet1;
quit;
data millet;
set p1(keep=coll);
run;
proc iml;
use ;
read all var _num_ into x;
use ;
read all var _num_ into y;
mt=j(1008,1,0);
do i=1 to 1008;
if y<=x[i]
then mt[i,1]=1;
end;
create a from mt;
append from mt;
quit;
run;
proc iml;
use a;
read all var _num_ into x;
s=sum(x)/1008;
end;
create ss from s;
append from s;
quit;
proc print data=ss;
run;

```

## I) Verifica della potenza dei test sotto $H_1$ e costruzione degli

### istogrammi

```

data nome file;
infile 'C:\Percorso\file.txt';
input f;
run;
proc iml;
use file;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.01
then mt[i,1]=1;
end;
create a1 from mt;
append from mt;
quit;
run;
proc iml;
use a1;
read all var _num_ into x;
s=sum(x)/1000;

```

```

create ss1 from s;
append from s;
quit;
proc print data=ss1;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.025
then mt[i,1]=1;
end;
create a2 from mt;
append from mt;
quit;
run;
proc iml;
use a2;
read all var _num_ into x;
s=sum(x)/1000;
create ss2 from s;
append from s;
quit;
proc print data=ss2;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.05
then mt[i,1]=1;
end;
create a3 from mt;
append from mt;
quit;
run;
proc iml;
use a3;
read all var _num_ into x;
s=sum(x)/1000;
create ss3 from s;
append from s;
quit;
proc print data=ss3;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.075
then mt[i,1]=1;
end;
create a4 from mt;
append from mt;
quit;

```

```

run;
proc iml;
use a4;
read all var _num_ into x;
s=sum(x)/1000;
create ss4 from s;
append from s;
quit;
proc print data=ss4;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.1
then mt[i,1]=1;
end;
create a5 from mt;
append from mt;
quit;
run;
proc iml;
use a5;
read all var _num_ into x;
s=sum(x)/1000;
create ss5 from s;
append from s;
quit;
proc print data=ss5;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.2
then mt[i,1]=1;
end;
create a6 from mt;
append from mt;
quit;
run;
proc iml;
use a6;
read all var _num_ into x;
s=sum(x)/1000;
create ss6 from s;
append from s;
quit;
proc print data=ss6;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.3

```

```

then mt[i,1]=1;
end;
create a7 from mt;
append from mt;
quit;
run;
proc iml;
use a7;
read all var _num_ into x;
s=sum(x)/1000;
create ss7 from s;
append from s;
quit;
proc print data=ss7;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.4
then mt[i,1]=1;
end;
create a8 from mt;
append from mt;
quit;
run;
proc iml;
use a8;
read all var _num_ into x;
s=sum(x)/1000;
create ss8 from s;
append from s;
quit;
proc print data=ss8;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.5
then mt[i,1]=1;
end;
create a9 from mt;
append from mt;
quit;
run;
proc iml;
use a9;
read all var _num_ into x;
s=sum(x)/1000;
create ss9 from s;
append from s;
quit;
proc print data=ss9;
run;
proc iml;

```

```

use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.6
then mt[i,1]=1;
end;
create a10 from mt;
append from mt;
quit;
run;
proc iml;
use a10;
read all var _num_ into x;
s=sum(x)/1000;
create ss10 from s;
append from s;
quit;
proc print data=ss10;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.7
then mt[i,1]=1;
end;
create a11 from mt;
append from mt;
quit;
run;
proc iml;
use a11;
read all var _num_ into x;
s=sum(x)/1000;
create ss11 from s;
append from s;
quit;
proc print data=ss11;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.8
then mt[i,1]=1;
end;
create a12 from mt;
append from mt;
quit;
run;
proc iml;
use a12;
read all var _num_ into x;
s=sum(x)/1000;
create ss12 from s;

```

```

append from s;
quit;
proc print data=ss12;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=0.9
then mt[i,1]=1;
end;
create a13 from mt;
append from mt;
quit;
run;
proc iml;
use a13;
read all var _num_ into x;
s=sum(x)/1000;
create ss13 from s;
append from s;
quit;
proc print data=ss13;
run;
proc iml;
use sp;
read all var _num_ into x;
mt=j(1000,1,0);
do i=1 to 1000;
if x[i]<=1
then mt[i,1]=1;
end;
create a14 from mt;
append from mt;
quit;
run;
proc iml;
use a14;
read all var _num_ into x;
s=sum(x)/1000;
create ss14 from s;
append from s;
quit;
proc print data=ss14;
run;proc iml;
use ss1;
read all var _num_ into s1;
use ss2;
read all var _num_ into s2;
use ss3;
read all var _num_ into s3;
use ss4;
read all var _num_ into s4;
use ss5;
read all var _num_ into s5;
use ss6;
read all var _num_ into s6;

```

```
use ss7;
read all var _num_ into s7;
use ss8;
read all var _num_ into s8;
use ss9;
read all var _num_ into s9;
use ss10;
read all var _num_ into s10;
use ss11;
read all var _num_ into s11;
use ss12;
read all var _num_ into s12;
use ss13;
read all var _num_ into s13;
use ss14;
read all var _num_ into s14;
mat=j(1,1,0);
f=mat//s1//s2//s3//s4//s5//s6//s7//s8//s9//s10//s11//s12//s13//s14;
create fi from f;
append from f;
quit;
proc print data=fi;
run;
```

## Bibliografia

- ✂ Addelman S. 1962, “**Orthogonal maineffect plans for asymmetrical factorial experiments**”, *Technometrics*, Vol. 4, No. 1, pp. 21– 46.
- ✂ Allenby G.M. and Ginter J.L. 1995, “**Using extremes to design products and segment markets**”, *Journal of Marketing Research*, Vol. 32, No. 4 (May), pp. 152–162.
- ✂ Allenby G.M., Arora N. and Ginter J.L. 1995, “**Incorporating prior knowledge into the analysis of conjoint studies**”, *Journal of Marketing Research*, Vol. 32, No. 2 (May), pp. 152–162.
- ✂ Arias J.T.G. 1996, “**Conjoint-based preferential segmentation in the design of a new financial service**”, *International Journal of Bank Marketing*, 14/3, pag 30-32,.
- ✂ Batsell R.R. and Lodish L.M. 1981, “**A model and measurement methodology for predicting individual consumer choice**”, *Journal of Marketing Research*, Vol. 18, No. 1 (February), pp. 1–12.
- ✂ Ben-Akiva M. and Gershensfeld S. 1998, “**Multi-featured products and services: Analyzing pricing and bundling strategies**”, *Journal of Forecasting*, Vol. 17, No. 3, pp. 175–196.
- ✂ Carmone F.J., Green P.E. and Jain A.K. 1978, “**The robustness of conjoint analysis: Some Monte Carlo results**”, *Journal of Marketing Research*, Vol. 15, No. 2, pp. 300–303.
- ✂ Carroll J.D. 1969, “**Categorical conjoint measurement**”, paper presented at the Annual Meeting of the Association of Mathematical Psychology, Ann Arbor, Michigan.
- ✂ Carroll J.D. and Green P.E. 1997, “**Psychometric methods in marketing research: Part 2, multidimensional scaling**”, *Journal of Marketing Research*, Vol. 34, No. 2 (May), pp. 193–204.
- ✂ Cattin P. and Wittink D.R. 1976, “**A Monte Carlo study of metricand nonmetric estimation techniques**”, Paper 341, Graduate School of Business, Stanford University.

- ✂ De Luca A. 2000, “**Le applicazioni dei metodi statistici alle analisi di mercato. Manuale di marketing quantitativo**”, Milano, F. Angeli.
- ✂ DeSarbo W.S., Wedel M., Vriens M. and Ramaswamy V. 1992, “**Latent class metric conjoint analysis**”, Marketing Letters, Vol. 3, No. 4 (July), pp. 273–288.
- ✂ Furio C., Sforza S. 2000, “**Indagini sulla soddisfazione del cliente: applicazioni della Conjoint Analysis**”, Roma.
- ✂ Gensch D.H. and Recker W.W. 1979, “**The multinomial multiattribute logitchoice model**”, Journal of Marketing Research, Vol. 16, No. 1 (February), pp. 124–132.
- ✂ Grandinetti R. 2002, “**Concetti e strumenti di marketing: il ruolo del marketing tra produzione e consumo**”, Milano, Etas.
- ✂ Green P.E. and Krieger A.M. 1991, “**Segmenting markets with conjoint analysis**”, Journal of Marketing, Vol. 55, No. 4 (October), pp. 20–31.
- ✂ Green P.E. and Krieger A.M. 1996, “**Individualized hybrid models for conjoint analysis**”, Management Science, Vol. 42, No. 6 (June), pp. 850–867.
- ✂ Green P.E. and Srinivasan V. 1978, “**Conjoint analysis in consumer research: issues and outlook**”, Journal of Consumer Research, Vol. 5, No. 2 (September), pp. 103–123.
- ✂ Green P.E. and Srinivasan V. 1990, “**Conjoint analysis in marketing: New developments with implications for research and practice**”, Journal of Marketing, Vol. 54, No. 1 (October), pp. 3–19.
- ✂ Green P.E. and Wind Y. 1973, “**Multiattribute Decisions in Marketing: A Measurement Approach**”, Dryden Press, Hinsdale, Illinois.
- ✂ Green P.E., Frank R.E. and Robinson P.J. 1967, “**Cluster analysis in test market selection**”, Management Science, Vol. 13, No. 8, pp. B387–400.
- ✂ Green P.E., Goldberg S.M. and Montemayor M. 1981, “**A hybrid utility estimation model for conjoint analysis**”, Journal of Marketing, Vol. 45, No. 1 (Winter), pp. 33–41.
- ✂ Gustaffsson A., Hermann A. and Huber F. 2000, “**Conjoint Measurement: Methods and Applications**”, Springer-Verlag, Berlin, Germany.

- ✎ Herman S. 1988, “**Software for full-profile conjoint analysis**”, in Proceedings of the Sawtooth Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, ed. M. Metegrano, Sawtooth Software, Ketchum, Idaho, pp. 117–130.
- ✎ Johnson R.M. 1974, “**Trade-off analysis of consumer values**”, *Journal of Marketing Research*, Vol. 11, No. 2 (May), pp. 121–127.
- ✎ Johnson R.M. 1987, “**Adaptive conjoint analysis**”, in Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing, ed. M. Metegrano, Sawtooth Software, Ketchum, Idaho, pp. 253–265.
- ✎ Kara A., Kaynak E., Kucukemiroglu O. 1996, “**An empirical investigation of US credit card users: card choice and usage behavior**”, *International Business Review* Vol.5, n°2, pag.209-230.
- ✎ Kara A., Kaynak E., Kucukemiroglu O. 1994, “**Credit card development strategies for youth market: the use of conjoint analysis**”, *International Journal of Bank Marketing*, vol 12, n°6, pp30-36.
- ✎ Kaul A. and Rao V.R. 1994, “**Research for product position and design decisions: An integrative review**”, *International Journal of Research on Marketing*, Vol. 12, No. 4, pp. 293–320.
- ✎ Kruskal J.B. 1965, “**Analysis of factorial experiments by estimating monotone transformations of the data**”, *Journal of the Royal Statistical Society, Series B*, Vol. 27, No. 2, pp. 251–263.
- ✎ Lenk P.J., DeSarbo W.S., Green P.E. and Young M.R. 1996, “**Hierarchical Bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs**”, *Marketing Science*, Vol. 15, No. 2, pp. 173–191.
- ✎ Luce R.D. and Tukey J.W. 1964, “**Simultaneous conjoint measurement: A new type of fundamental measurement**”, *Journal of Mathematical Psychology*, Vol. 1, pp. 1–27.
- ✎ Mankila M. 2003, “**Retaining students in retail banking through price bundling: Evidence from the Swedish market**”, *European Journal of Operational Research*.
- ✎ McFadden D. 1974, “**Conditional logit analysis of qualitative choice behavior**”, in *Frontiers on Econometrics*, ed. P. Zarembka, Academic Press, New York, pp. 105–421.

- ✂ Moore W.L. 1980, “**Levels of Aggregation in Conjoint Analysis: An Empirical Comparison**”, Journal of Marketing Research, (Nov.).
- ✂ Murphy M.C., Soutar G.N. 2003, “**What individual investor value: Some Austalian evidence**”, Journal of Economic Psychology.
- ✂ Muzyka D., Birley S., Leleux B. 1996, “**Trade-offs in the investment decisions of european venture capitalists**”, Journal of Business Venturing 11, pag.273-287.
- ✂ Nair S.K., Tarasewich P. 2003, “**A model and solution method for multi-period sales promotion design**”, European Journal of Operational Research, n°150, pag 672-687.
- ✂ Oppewal H., Vries M. 2000, “**Measuring perceived service quality using integrated conjoint experiments**”, International Journal of Bank Marketing, 18/4, pag.154-169.
- ✂ Pesarin F., Lago A. 1998, “**Nonparametric combination of dependent rankings with application to the quality assessment of industrial products**, Padova, Univ. di Padova, Dip. di Scienze Statistiche.
- ✂ Punj G.N. and Staelin R. 1978, “**The choice process for graduate businessschools**”, Journal of Marketing Research, Vol. 15, No. 4 (November), pp. 588–598.
- ✂ Ramaswamy V. and Cohen S.H. 2000, “**Latent class models for conjoint analysis**”, in Conjoint Measurement: Methods and Applications, eds. A. Gustafsson, A. Hermann, and F. Huber, Springer-Verlag, Berlin, Germany.
- ✂ Saaty T.L. 1980, “**The Analytical Hierarchy Process**”, McGraw-Hill, New York.Chesnais
- ✂ Sawtooth Software 1999, *CBC for Windows*, Sequim, Washington.
- ✂ Shepherd D.A. 1999, “**Venture capitalist’s assessment of new venture survival**”, Management Science Vol 45, n° 5 (May).
- ✂ Shocker A.D. and Srinivasan V. 1977, “**LINMAP (Version II): A FORTRAN IV computer program for analyzing ordinal preference (dominance) judgments via linear programming techniques and for conjoint measurement**”, Journal of Marketing Research, Vol. 14, No. 1, pp. 101–103.

- ✎ Srinivasan V. 1988, “**A conjunctive compensatory approach to the selfexplication of multiattributed preferences**”, *Decision Sciences*, Vol. 19, No. 2 (Spring), pp. 295–305.
- ✎ Talaga J.A., Buch J. 1998, “**Consumer trade-offs among mortgage instrument variables**”, *International Journal of Bank Marketing*, pag-264-270, 16/6.
- ✎ Vavra T.G., Green P.E. and Krieger A.M. 1999, “**Evaluating EZ-Pass**”, *Marketing Research*, Vol. 11, No. 3 (Summer), pp. 5–16.
- ✎ Westwood D.L.T. and Beazley D. 1974, “**The trade-off model and its extensions**”, *Journal of the Market Research Society*, Vol. 16, No. 3, pp. 227–241.
- ✎ Wind J., Green P.E., Shifflet D. and Scarbrough M. 1989, “**Courtyard by Marriott: Designing a hotel facility with consumer-based marketing models**”, *Interfaces*, Vol. 19, No. 1 (January-February), pp. 25–47.
- ✎ Wittink D. and Cattin P. 1989, “**Commercial use of conjoint analysis: An update**”, *Journal of Marketing*, Vol. 53, No. 3 (July), pp. 91–96.
- ✎ Wittink D., Vriens M. and Burhenne W. 1994, “**Commercial use of conjoint in Europe: Results and critical re.ections**”, *International Journal of Research in Marketing*, Vol. 11, N. 1, pp. 41–52.
- ✎ Young F.W. 1969, “**Polynomial conjoint analysis of similarities: De.nitions for a special algorithm**”, Research paper No. 76, Psychometric Laboratory, University of North Carolina.
- ✎ Zani S. 1989, “**Classificazioni “sfocate” di unità territoriali: un’applicazione alle regioni italiane**”, *Atti delle Giornate di studio su Analisi Statistica di Dati Territoriali*, pp.495-506.