



UNIVERSITÀ DEGLI STUDI DI SASSARI



DIPARTIMENTO DI ARCHITETTURA, DESIGN E URBANISTICA

Corso di Dottorato in Architettura e Ambiente

XXIX Ciclo

Camminabilità, perché e come misurarla:

un modello e uno studio empirico di validazione della misura

RELATORE

Prof. Ivan Blečić

DOTTORANDO

Dario Canu

2017

La giustizia non riguarda soltanto
il modo giusto di distribuire le cose,
ma il modo giusto di valutarle.

Michael J. Sandel

INDICE

Riassunto.....	7
Introduzione.....	8
Il diritto alla città	13
1.1 <i>Il diritto alla città di Lefebvre.....</i>	<i>15</i>
1.2 <i>Il diritto alla camminabilità.....</i>	<i>17</i>
1.3 <i>L'approccio delle capacità di Sen.....</i>	<i>20</i>
1.4 <i>Sul capability approach, applicazioni.....</i>	<i>22</i>
Come e cosa misurare (dall'accessibilità alla camminabilità)	26
2.1 <i>Concetti di Equità.....</i>	<i>27</i>
2.2 <i>Definizioni di equità spaziale.....</i>	<i>29</i>
2.3 <i>Definizioni di accessibilità.....</i>	<i>31</i>
2.4 <i>Differenze tra Accessibilità, Accesso, e Camminabilità.....</i>	<i>33</i>
2.5 <i>Metodi di misura dell'accessibilità.....</i>	<i>36</i>
2.5.1 <i>Metodi di misurazione, Place-Based Measures</i>	<i>39</i>
2.5.2 <i>Metodi di misurazione, People-Based Measures</i>	<i>43</i>
2.5.3 <i>Alcuni approfondimenti sul Modello Network Analysis</i>	<i>45</i>
2.6 <i>Discussione</i>	<i>47</i>
La misura della camminabilità.....	49
3.1 <i>Gli elementi della misura di camminabilità.....</i>	<i>49</i>
3.1.1 <i>Alcuni strumenti di misura</i>	<i>52</i>
3.2 <i>Una misura della camminabilità.....</i>	<i>56</i>

3.3	<i>Percezione e caratteristiche dello spazio</i>	61
3.3.1	<i>Variabili e domande</i>	63
3.3.2	<i>Metodi di campionamento</i>	66
3.3.3	<i>Test simili, metodi di analisi e risultati</i>	69
	Test empirici e modelli predittivi	77
4.1	<i>Segmenti stradali</i>	78
4.1.1	<i>Costruzione dell'esperimento</i>	79
4.1.1.1	<i>Regressione lineare multipla</i>	82
4.1.2	<i>Alcuni altri modelli di regressione e analisi</i>	87
4.1.2.1	<i>Regressione stepwise</i>	87
4.1.2.2	<i>Regressione logistica ordinale</i>	88
4.1.2.3	<i>Test di modelli non lineari</i>	94
4.1.2.4	<i>Dipendenza tra le variabili</i>	95
4.1.2.5	<i>Relazione tra le variabili, il path diagram</i>	97
4.1.3	<i>Compensazione</i>	101
4.1.3.1	<i>Dominance-based rough set</i>	103
4.1.3.2	<i>Ordinamento per regola di maggioranza</i>	105
4.1.3.3	<i>Mr-Sort + rough set = ELECTRE TRI</i>	108
4.1.3.4	<i>Reti neurali artificiali</i>	109
4.1.4	<i>Alcune considerazioni sul confronto tra i diversi metodi</i>	112
4.2	<i>Attraversamenti alle intersezioni</i>	113
4.2.1	<i>Costruzione dell'esperimento</i>	114
4.2.1.1	<i>Regressione lineare multipla</i>	118
4.2.2	<i>Alcuni altri modelli di regressione</i>	124
4.2.2.1	<i>Regressione stepwise</i>	125
4.2.2.2	<i>Regressione logistica ordinale</i>	126

4.2.2.3 Dipendenza tra le variabili.....	128
4.2.3 Compensazione.....	129
4.2.3.1 Dominance-based rough set	130
4.2.3.2 Ordinamento per regola di maggioranza	130
4.2.3.3 Mr-Sort + rough set = ELECTRE TRI	133
4.2.3.4 Reti neurali artificiali	134
4.3 Discussione sui modelli di regressione e stima dei giudizi	135
4.3.1 Il problema dell'overfitting.....	139
4.3.2 Rete neurale, k-fold Cross-Validation	141
4.3.3 Il benchmark.....	143
4.4 Trade-off tra segmenti, attraversamenti e distanze.....	150
4.5 Possibili usi dei dati e dei modelli	152
Conclusioni.....	155
Appendice A: I modelli ad equazione strutturale.....	160
Appendice B: Il rough set approach.....	170
Appendice C: I modelli Majority Rule Sorting ed ELECTRE TRI.....	190
Appendice D: Le reti neurali.....	197
Bibliografia	204

Riassunto

Il contenuto di questo testo riflette sugli aspetti dell'accessibilità pedonale nello spazio urbano, di come questa sia un influente aspetto di giustizia e su come la camminabilità possa essere misurata e modellizzata. Il prodotto finale della ricerca può essere suddiviso in tre grandi aree:

La prima, introduttiva, ragiona sugli aspetti dell'accessibilità pedonale e delle sue implicazioni nella giustizia urbana. Si propone uno dei possibili punti di vista per cui una città camminabile è un diritto e lo si fa secondo la teoria del diritto alla città e dell'approccio delle capacità.

La seconda, di ricerca bibliografica, pone in rassegna i metodi di misura dell'accessibilità con particolare attenzione alla sua definizione rispetto alla giustizia e all'equità. L'accessibilità intesa come misura della giustizia urbana è dipendente da cosa si intende per *equità*, per *equità urbana* e per *accessibilità*. Ogni diversa concezione delle definizioni porta ad un approccio alla misura differente, basato sulle caratteristiche dello spazio o della persona. Questo lavoro ha il pregio di distaccarsi dalla letteratura ed affrontare la misura includendo caratteristiche di entrambi gli approcci.

La terza, finale, si compone di due esperimenti empirici per definire e validare dei modelli di classificazione, integrando un'ampia raccolta bibliografica di ricerche simili ed una serie di test sui modelli. Il risultato delle prove sperimentali consente di determinare quali sono le caratteristiche dello spazio percepite come importanti nella valutazione dell'ambiente urbano e qual è il loro peso nel giudizio.

A partire dai risultati dei test di regressione si propongono alcuni strumenti di classificazione capaci di inferire parametri e struttura dal campione di dati per la classificazione di nuove alternative. I risultati consentono di definire il miglior modello di ordinamento, a partire dal quale è possibile costruire uno strumento di valutazione della camminabilità.

Introduzione

Questo testo parla di camminabilità! Una strada è una strada, ma c'è strada e strada. Quindi potremmo chiederci, qual è una buona strada camminabile? Prendere una decisione è difficile, tutto dipende dalle caratteristiche dello spazio, da cosa ci interessa e da quanto dobbiamo camminare.

Sulle caratteristiche dello spazio, nella lingua italiana, la denominazione urbanistica generica ci dà una certa indicazione: via, vicolo, viale, corso, indicano tutti un certo tipo di strada ma con caratteristiche differenti.

Ma se ci serviamo dell'inglese americano, autologico, la nomenclatura è più dettagliata ed esprime meglio le proprietà dello spazio pubblico. Una strada generica, una *road*, indica tutto ciò che connette due punti, una *street* è invece una strada pubblica urbana con edifici su entrambi i lati e qualche albero. Poi la *avenue*, perpendicolare alle *street*, sempre con edifici o alberi ma più larga e con una velocità maggiore, di solito termina con un edificio civile o un monumento. Il *boulevard* ha invece sempre gli alberi su entrambi i lati e ha spesso un separatore di traffico centrale, è generalmente più lunga della *avenue*. Il loro opposto è una *lane*, una strada stretta in zona periferica, ha una pavimentazione di bassa qualità ma è un ottimo spazio sicuro di gioco per i bambini. Poi c'è il *drive*, una strada tortuosa che segue le forme dell'ambiente che circonda, come un lago o un fiume, se fiancheggia la costa prende il nome di *esplanade* e quando è principalmente pedonale diventa una *promenade*. Una *place* è una strada senza uscita, una *court* una strada senza uscita a *cul-de-sac*. *Alley* indica invece il percorso di accesso per la manutenzione sul retro degli edifici, senza marciapiedi o negozi, è simile al *passage*, molto stretto ma che serve per attraversare lunghi blocchi o per collegare parcheggi e strada principale, è spesso coperto e con vetrine di negozi. Infine il *path*, una via pedonale o ciclabile, di connessione, che attraversa un parco.

Ma allora, è meglio il *boulevard*, alberato, ricco di servizi ma molto trafficato, o la *lane*, sicura ma periferica? è meglio la *alley* o la *street*? quale tra queste è una buona strada per camminare? Ancora non lo sappiamo, ci serve leggere questo testo! Il suo scopo è quello di identificare quali sono gli elementi dello spazio che lo rendono pedonabile, ci serve per misurare il grado di camminabilità dello spazio.

Ma per quale motivo una città deve essere camminabile? Ci sono diversi motivi: per incentivare l'esercizio fisico, per la socialità, per lo sviluppo economico, per ridurre l'inquinamento e per promuovere il concetto – che può essere letto nel conflitto tra pedoni e veicoli – del diritto alla città.

Se immaginiamo una tipica strada urbana e prestiamo attenzione allo spazio, la maggior parte è dedicato alle auto, non ne avanza molto altro. Ciò che rimane sono stretti marciapiedi ai lati delle strade carrabili e qualche attraversamento qua e là per connetterli. Le auto dominano la città!

La presenza di così tanti veicoli nelle strade comporta una pluralità di limiti alla camminabilità, non solo in termini di conflitto dello spazio, ma anche di salute, tempo, costo sociale, costo ambientale, sicurezza, ecc.

In molte città la riappropriazione dello spazio urbano è un *mainstream*, la città più virtuosa al momento è forse Barcellona, con il suo progetto delle *superilles* che restituisce lo spazio ai quartieri e riduce l'inquinamento cittadino.

Ma non possiamo limitare il traffico in tutte le vie, questa non è una soluzione. Allora ciò che ci interessa fare, ciò che tenta di fare questo testo, è misurare quanto lo spazio sia camminabile per offrire uno strumento di supporto alla decisione, di gestione dello spazio, di scelte localizzative, di progettazione.

Perciò in questa tesi si affronta uno tra i diversi motivi per cui una città debba essere camminabile, da uno dei possibili punti di vista, quello del rapporto tra pedoni e veicoli. Si vuole giustificare la promozione dell'accessibilità secondo la giustizia urbana e – tra le possibili teorie – le tesi del diritto alla città di Lefebvre e dell'approccio alle capacità di Sen sono le preferibili.

La teoria del diritto alla città di Lefebvre è la teoria di base da cui derivano (in maniera diretta o indiretta) i successivi approcci di giustizia urbana (ad esempio quelli di Harvey, Soja e Fainsten), è fondata sul concetto di valore d'uso (città come *opera*) e valore di scambio (città come *prodotto*) che può essere paragonato all'uso dello spazio di pedoni e veicoli. I primi usano la città, ne producono le condizioni e la vita sociale, gli altri la sfruttano e la consumano creando svantaggi e limiti al pedone e (in maniera diretta e indiretta) generano una città a misura di *abitante-tipo* (adulto, maschio, sano, auto-munito). Per questo, in parallelo alla teoria di Lefebvre, si può dire che il diritto alla città per i pedoni è corrotto o – in alcuni casi – negato.

La teoria delle capacità di Sen fornisce un doppio supporto al lavoro, teorico e metodologico. Dal punto di vista teorico può giustificare l'importanza della mobilità lenta per due ragioni: la prima è che dobbiamo promuovere una città dove gli individui – a prescindere dai funzionamenti – abbiano le stesse capacità di movimento, siano perciò tutti messi in condizioni di equità e di pari libertà. La seconda è che alcune di queste capacità sono negate da comportamenti sociali – figli dell'automobile, come la diffusione urbana – che creano situazioni nelle quali l'*abitante-non-tipo* subisce una privazione delle possibilità di movimento; chi si trova in situazioni di svantaggio deve poter essere messo in condizioni di parità. Dal punto di vista metodologico, la teoria delle capacità getta le basi per una misura dell'accessibilità che oltre alle componenti dello spazio a disposizione degli individui ne misura la capacità e l'interesse d'uso.

L'approccio alla misura della camminabilità non è quindi limitato alle sole componenti dello spazio, ma tiene conto delle caratteristiche individuali, in due modi: indagando le percezioni e costruendo un metodo di valutazione che non

contabilizzi le disposizioni urbane (*commodities*) ma tenga conto della capacità di fruirne (*capabilities*).

Per ottenere una misura formale della camminabilità ciò che serve è una *proxy*, un elemento che sia misurabile. Nella letteratura di riferimento quando si parla di spostamenti nello spazio c'è un concetto che è comunemente condiviso, l'indicatore della capacità di movimento, l'accessibilità. Ma in questo contesto di accessibilità come elemento di giustizia urbana, la sua misura formale è dipendente dall'approccio ai concetti di equità, equità urbana ed accessibilità stessa. Una profonda ricerca bibliografica sul tema porta a sistematizzare le diverse definizioni e definire i possibili percorsi logici che portano alla misura formale di accessibilità, che, a seconda di come si intendano le tre definizioni elementari, porta a due tipi di misure base: una basata sulle caratteristiche dello spazio (*place-based*) ed una basata sulle caratteristiche degli individui (*people-based*). L'idea di fondo del lavoro di ricerca, in accordo con gli scopi e con gli approcci teorici, è di interpretare l'accessibilità pedonale a partire dalle caratteristiche dello spazio ma secondo una prospettiva soggettiva, in particolare di percezione di qualità dello spazio.

La percezione di qualità dello spazio ci permette di interpretare bisogni, necessità e desideri (gli aspetti della teoria delle capacità) in funzione di come gli individui valutano l'ambiente che li circonda; questo è l'elemento che, in relazione alle caratteristiche dello spazio, fornisce le basi della misura. Questo punto, il nocciolo della ricerca, si interessa quindi di testare empiricamente la percezione di qualità degli individui e la pone in relazione agli attributi dello spazio per rispondere ai due obiettivi: (1) identificare le caratteristiche urbane avvertite come importanti nella valutazione di camminabilità (attraverso gli strumenti della statistica inferenziale) e (2) modellarle in un sistema di classificazione automatico, validato e calibrato sui dati raccolti.

La qualità dello spazio è indagata con gli strumenti della psicometria, mentre i modelli di previsione sono costruiti con le tecniche di supporto alla decisione dell'analisi multicriteriale e di *supervised machine learning*.

Lo studio si concentra sulla percezione dell'ambiente pedonale dei segmenti stradali (tratti omogenei di strada) e delle intersezioni (punto di interazione tra il pedone e l'auto) in una prospettiva nuova rispetto alla letteratura. In confronto agli studi simili e rispetto ai test disponibili in letteratura questo lavoro ha il vantaggio di considerare un numero di aspetti dello spazio (variabili) ed un dettaglio (modalità) maggiore. A differenza degli studi citati, la misura della camminabilità include inoltre gli attraversamenti alle intersezioni; questi sono un importante elemento di disturbo della camminabilità e, di conseguenza, influenzano fortemente la sua valutazione.

In ogni esperimento un gruppo di intervistati ha percorso i segmenti e gli attraversamenti raccogliendo dati sulle caratteristiche dello spazio (variabili indipendenti) e sulla loro percezione di qualità (variabile dipendente). Dallo studio è possibile inferire gli attributi dell'ambiente urbano percepiti come importanti dagli intervistati, rispondendo così al primo obiettivo dell'esperimento. Attraverso le procedure di regressione sono risultati significativi nove attributi per i segmenti e dieci per gli attraversamenti, confermando ed ampliando gli studi della letteratura.

Il secondo obiettivo dell'esperimento è quello di definire un sistema di valutazione della qualità dello spazio pedonale a partire dalle sue caratteristiche. A questo scopo sono stati comparati diversi modelli per individuare il più adatto alla misura degli aspetti studiati. Nello specifico i modelli testati possono essere classificati in due grandi famiglie, metodi compensativi e metodi non-compensativi. I primi, in questo caso basati su diverse metodologie di regressione (lineare multipla, *stepwise* e ordinale), stimano le classificazioni con funzioni a somma pesata; i secondi stimano i giudizi dalla definizione di regole di assegnazione, soglie di maggioranza e percorsi di classificazione. Il vantaggio dei metodi non-compensativi è di gestire alcuni aspetti di dipendenza e covarianza delle variabili (indagati con test collaterali) e – più importante in questo caso – permettono di ottenere classificazioni più accurate.

Nello specifico i metodi non-compensativi sono: il *dominance-based rough set approach* che, a partire dal set di dati, estrae regole di assegnazione del tipo *if... then*; il metodo *Majority Rule Sort* (e ELECTRE TRI) che classifica, secondo una soglia di maggioranza, le alternative in base all'appartenenza di coalizioni di criteri ad una classe; le reti neurali artificiali che, simulando il funzionamento del cervello, consentono di classificare le alternative attraverso percorsi non lineari composti da neuroni e connessioni inter-neurali.

Ognuno dei modelli proposti ha differenti vantaggi e caratteristiche desiderabili per definire un modello di previsione della camminabilità urbana. La scelta di quale tra questi è il migliore è condotta valutando la percentuale di alternative classificate correttamente ed in funzione di qual è il livello di sovra-adattamento del modello.

Le reti neurali artificiali si attestano come il miglior modello che permette di classificare correttamente il 60-70% dei dati, che – nel campo delle scienze sociali – può essere ritenuto un buon risultato.

Il modello ottenuto può essere utilizzato come strumento di supporto alla decisione nelle scelte di mobilità urbana o può essere integrato in modelli di gestione della città o valutazione della camminabilità. Può essere applicato per disegnare mappe di accessibilità e/o può essere integrato in strumenti più complessi come strumenti di navigazione, scelte localizzative, valutazione della qualità urbana o di altri aspetti urbani.

Risponde all'obiettivo principe di questa ricerca – la valutazione dell'equità urbana intesa come capacità di movimento a piedi nello spazio – permettendo di definire mappe di “accessibilità minima” o accessibilità in funzione delle caratteristiche di ogni individuo, fornendo uno strumento di confronto tra classi di soggetti e tra città o quartieri.

Ma può essere applicato anche come uno strumento di supporto alla decisione per le amministrazioni, integrato in strumenti partecipativi e inclusivi di raccolta dei dati collettivi o di valutazione dei progetti di modifica dello spazio pedonale. Può essere uno strumento di supporto alle scelte localizzative nella città, di complemento alla stima dei valori immobiliari o può essere uno strumento di navigazione pedonale (un *google maps* per pedoni) che tenga conto degli elementi importanti per la camminabilità, non limitando la misura alle sole distanze.

Capitolo 1

Il diritto alla città

Sappiamo bene che le città nascono perché è vantaggioso aggregarsi in comunità, la prima città ritenuta tale, Uruk¹, nell'antica Mesopotamia, già dal quarto millennio a.C. contava tra i 50 e gli 80 mila abitanti², distribuiti su una superficie di 6 km², con una densità (~13.000 ab/km²) da fare invidia alla moderna New York³ (~ 10.000 ab/km²). Teniamo a mente questi numeri, non sono qui per caricare il testo, servono a farci riflettere sull'argomento centrale di questo lavoro, la capacità, la possibilità, l'abilità, la facoltà di muoversi, a piedi, nella città⁴.

Le città moderne, nelle quali oggi viviamo, sono state profondamente modificate dalla rivoluzione industriale, quando la nascita delle fabbriche ha generato i movimenti di lavoratori dalle campagne verso le malsane periferie (spesso non pianificate) delle città rappresentate da Gustave Doré⁵. Ne parliamo perché, la rivoluzione industriale, ci interessa per due motivi: il primo è che getta le fondamenta dell'urbanistica moderna, ci interessa perché influenzò la forma delle città nelle quali abitiamo, ci interessa ragionare sulle conseguenze della vita sociale e sui comportamenti che ne derivano. Il secondo motivo è che proprio a causa dello sviluppo della città capitalistica⁶, conseguenza delle rivoluzioni industriali, nascono gli scritti di Henri Lefebvre sul diritto alla città, che stanno alla base di questo lavoro.

¹ Si veda ad esempio: Liverani M. (2006) *Uruk la prima città*, Laterza.

² Si veda ad esempio: Harmansah, Ö. (2007). *The Archaeology of Mesopotamia: Ceremonial centers, urbanization and state formation in Southern Mesopotamia*.

³ Uruk vs New York: ~13.000 ab/km² vs ~ 10.000 ab/km².

⁴ Non a caso, densità superiori a Uruk sono oggi misurabili principalmente in India (Mumbai supera 31 mila ab/km²), dove il numero di auto è 20,8 milioni (2010) per 1,25 miliardi di abitanti (2013), molte meno rispetto ai 239,8 milioni di veicoli negli USA che contano 318,9 milioni di abitanti (2014).

⁵ Il riferimento è ad esempio a: *Over London by Rail*, Gustave Doré, 1870.

⁶ Si veda ad esempio: Harvey D. (2012) *Il capitalismo contro il diritto alla città*, Ombre Corte.

Uno degli aspetti della crescita delle città, in concomitanza con la nascita e diffusione dell'automobile (figlia della rivoluzione industriale⁷), sta proprio nell'influenza che la mobilità veloce ha avuto – ed ha – nello sviluppo delle nostre città contemporanee, «ogni modo di produzione, ha avuto la propria tipologia di città» scrive Lefebvre (2014, p.123).

È facile intuire che vivere in comunità crea svantaggi, problemi e conflitti; nella città moderna, più ci si avvicina al centro più aumentano i prezzi degli alloggi, aumenta l'inquinamento, il rumore, la congestione, ecc. Ed ecco che l'auto consente di superare i limiti del *vivere in città* pur premettendo di *vivere la città* grazie alla facilità e rapidità di movimento. Si verifica quindi l'inversione di tendenza dei movimenti dalla campagna alla città del '700, con la fusione tra città e campagna (il neologismo “rurbano” citato e descritto da Lefebvre (2014, p.74)) e con l'urbanizzazione dell'agro.

Ed ecco che torna utile ripensare ai numeri di Uruk e New York. Possiamo supporre che l'alta densità della prima città, oltre che per impressionare a scopo difensivo (Filupucci 2015, p.162), fosse causa della necessità di rendere la città camminabile a piedi, mentre New York, ma più in generale le città moderne, hanno potuto espandersi sino ai sobborghi anche grazie ai motoveicoli. Ma lasciando da parte le grandi metropoli americane, che richiederebbero sforzi di analisi più profondi, possiamo guardare alle città medio-grandi europee, che assoggettate dal fenomeno della città diffusa o dei territori urbanizzati⁸ soffrono dei limiti alla camminabilità che vogliamo trattare.

Ci chiediamo quindi se l'automobile non abbia giocato un ruolo fondamentale influenzando la città che oggi è spesso costruita, disegnata, organizzata o autoprodotta⁹ secondo le necessità degli automobilisti e non dei pedoni.

Nella città esiste un conflitto di esigenze e spazi tra pedoni e automobilisti, Cecchini e Talu (2011) sostengono che «poiché esiste un conflitto pressoché

⁷ Il primo prototipo di auto con motore a vapore fu il carro di Cugnot progettato nel 1769.

⁸ Si veda ad esempio: Indovina F. (2009) *Dalla città diffusa all'arcipelago metropolitano*, Franco Angeli.

⁹ Ad esempio la *città diffusa*, si genera spontaneamente, in maniera non pianificata, a partire da comportamenti sociali.

irriducibile fra le esigenze degli automobilisti e quelle dei pedoni, si può affermare che la città, di fatto, non è capace di garantire ad una parte dei suoi abitanti la possibilità di esercitare pienamente un diritto fondamentale, quello di muoversi liberamente e autonomamente». Esiste più d'un conflitto tra pedoni e automobilisti, l'auto genera congestione, crea confusione, riduce l'efficacia del trasporto pubblico, riduce la sicurezza dei pedoni, produce inquinamento atmosferico, acustico ed estetico, consuma il suolo riducendo lo spazio per i pedoni. L'auto genera perciò una situazione di iniquità tra l'automobilista e il pedone, quest'ultimo è succube degli effetti negative generati dalle auto pur non contribuendo a generali. Dal punto di vista di questa tesi, l'urbanistica deve promuovere l'uguaglianza tra i cittadini fornendo a tutti gli abitanti gli stessi diritti di accesso ed uso dello spazio, dei suoi servizi e delle opportunità urbane. Compito degli urbanisti è perciò costruire una città che garantisca il diritto di muoversi per tutti, attraverso progetti e politiche d'incentivo alla mobilità dolce e non motorizzata¹⁰. Far sì che la città garantisca a tutti gli abitanti gli stessi diritti di accesso ed uso di spazi, servizi e opportunità urbane è forse l'obiettivo principale dell'urbanistica. L'argomento centrale di questo testo è quindi la capacità di camminare a piedi nella città. Possiamo leggerlo come un diritto fondamentale (poi diremo perché) e differenziarlo da accesso e accessibilità. La riflessione in questo testo nasce guardando agli scritti di Lefebvre sul diritto alla città; come negli anni '60 descritti dall'autore, la città vive una seconda crisi che può essere letta (anche) nella diffusione urbana e nelle difficoltà di camminare nella città.

1.1 Il diritto alla città di Lefebvre

Possiamo ora far ruotare il discorso attorno a due concetti chiave del pensiero di Lefebvre, che possono essere applicati in modo parallelo a ciò che trattiamo in questo testo. Parliamo della dicotomia tra il valore d'uso (la città e la vita urbana, il tempo urbano) e il valore di scambio (gli spazi acquistati e venduti, il consumo dei prodotti, dei beni, dei luoghi e dei segni) che è per Lefebvre un punto cardine

¹⁰ Questo non significa solo rendere svantaggioso l'uso dell'auto, ma piuttosto fornire delle valide alternative agli automobilisti perché possano essere più pedoni che autisti.

della crisi della città, che mette in opposizione la città come *opera* e come *prodotto*. La città intesa come opera è il risultato della vita urbana, delle interazioni tra gli utilizzatori dello spazio, è il “luogo della festa”, generato, modificato, sviluppato, dalla comunità e dalla collettività. In questo senso la città ha un valore d’uso, la sua ricchezza è costituita dalla possibilità degli abitanti di utilizzare uno spazio comune e comunemente costruito. Dall’altro lato si contrappone il suo valore di scambio. Secondo Lefebvre, con l’industrializzazione (già nel XIX secolo) inizia un processo di mercificazione della città – che verrà poi individuato da David Harvey nei processi del capitalismo moderno¹¹ – che guarda alla città in un’ottica di consumo e mercificazione, individuando in essa un elemento di profitto, da usare, commerciare, produrre e scambiare. Con il radicarsi dal capitalismo e del processo di industrializzazione, il valore di scambio della città come “oggetto” si sostituisce al suo valore d’uso, privando i cittadini del senso fondamentale della vita urbana della città come luogo della partecipazione, dell’incontro e dell’interazione (Lefebvre, 2014, p.8).

La più famosa tra le frasi di Lefebvre è forse «Spazio (sociale) è un prodotto (sociale)» (Lefebvre, 1974). Questa locuzione sintetizza in maniera estremamente efficace gli elementi chiave del concetto del diritto alla città: innanzitutto la città non è solo un luogo dove si *svolgono* le relazioni sociali, ma nella città si *producono* le relazioni sociali che a loro volta producono la città, i comportamenti plasmano la forma della città e la forma della città influenza i comportamenti e le relazioni. Lo spazio è quindi sociale, è il luogo delle interazioni, che forma – ed è formato – dai comportamenti. Lo spazio è però un prodotto, Lefebvre non nega che la produzione industriale, dopo un certo livello di crescita, genera urbanizzazione, ne fornisce le condizioni, ne stabilisce le possibilità. Ma il prodotto è anche sociale, la città è generata dai processi sociali, dalle interazioni operate da un processo di sviluppo sociale che possiede la paternità dello sviluppo della città, che è vista come una «proiezione della società sul territorio» (Lefebvre, 2014, p.63). “Lo spazio sociale è un prodotto sociale” sta quindi ad indicare che lo sviluppo dello spazio, che si genera dai processi sociali,

¹¹ Si veda ad esempio: Harvey 2012, *Il capitalismo contro il diritto alla città*, Ombre Corte.

è il risultato della vita dei cittadini. Se è la società che è causa della produzione della città allora questa ha il diritto di decidere il suo sviluppo, limitando l'intervento del processo produttivo nello sviluppo della città.

Il diritto alla città di Lefebvre si fonda su questo aspetto, sull'idea che la città sia un prodotto della comunità che ha il diritto di decidere dei – e usare i – propri spazi. In sintesi, il diritto alla città è inteso come un diritto «alla vita urbana, alla centralità ritrovata, ai luoghi d'incontro e scambio, ai ritmi di vita e ai modi di utilizzare il tempo che consentano un *uso* pieno e completo di momenti, luoghi, ecc.» (Lefebvre, 2014, p.136).

1.2 Il diritto alla camminabilità

Abbiamo anticipato che la teoria di Lefebvre si sviluppa alla fine degli anni '60, analizzando il diritto alla città della classe operaia sulla classe borghese e industriale; la teoria dell'autore è stata negli anni adattata a differenti situazioni e riscopre oggi una nuova necessità di applicazione. Oggi si legge il diritto alla città nei movimenti sociali di riappropriazione dello spazio, come nelle lotte a *Gezi Park* (approfonditamente descritti da David Harvey¹²) o nelle rivolte del 2005 delle *balineue* francesi, nella *People's park protest* del Maggio 1969 in California o nelle più contemporanee azioni di *Urbanismo Tattico*¹³; lo si riscopre nelle esigenze collettive di un ambiente sano, di migliori condizioni di vita (servizi, trasporti, istruzione), di accesso alla casa per tutti, ma anche nel diritto dei beni pubblici (l'acqua, l'ambiente, ecc.) e nel diritto di cittadinanza (la partecipazione, il coinvolgimento politico, ecc.). Per ultimo, ma non meno importante, il diritto alla città è oggi – esattamente come lo era per Lefebvre cinquanta anni fa – un diritto di rivalsa dell'uso e della gestione pubblica e comune dello spazio, in opposizione alle dinamiche della proprietà e del diritto edificatorio, dove la città serve le esigenze degli investitori, dei costruttori, degli speculatori e non gli interessi della collettività.

¹² Si veda ad esempio: Harvey, D. (2012). *Rebel cities. from the right to the city to the right to the urban revolution*. Verso.

¹³ Si veda ad esempio: Lyon M, Douglas, G. (2015). *Tactical Urbanism: Short-term Action for Long-term Change*, Island press.

Nello specifico del nostro caso, abbiamo detto che esiste un conflitto tra auto e pedone, abbiamo detto che la prima influenza la forma della città (mettendola in crisi) a discapito di chi percorre lo spazio a piedi, aumentando le distanze (se non causa è perlomeno complice della diffusione urbana!), riducendo lo spazio per i pedoni e incrementando i livelli di inquinamento, pericolo, ecc. Ebbene, non si tratta forse un'analogia con la dicotomia tra collettività e industrializzazione descritti da Lefebvre (o collettività e capitalismo raccontati da Harvey)? Pare di sì, possiamo allora sostenere che ci sia una qualche similitudine tra *opera e prodotto* e tra pedoni e automobilisti nell'uso dello spazio. Se leggiamo all'uso che queste due categorie fanno della città possiamo sostenere che i pedoni producono la vita sociale nella città, si muovono, si incontrano, giocano, stanno seduti, vendono e comprano. Mentre gli automobilisti – quando si spostano in auto – usano la città solo come mezzo di comunicazione tra punti, occupano lo spazio senza apparenti vantaggi per la comunità, creano traffico, inquinamento e pericolo; in quel momento stanno leggendo – della città – solo il valore di scambio (inteso come consumo dei luoghi). C'è da aggiungere (e lo vedremo tra poco) che possiamo avere una visione della città come *prodotto* anche nella gestione della città da parte di chi la disegna, che guarda alla strada (nuovo spazio pubblico, che sostituisce la piazza) come uno strumento per il passaggio e la sosta dei veicoli. Possiamo perciò intendere il diritto alla città anche come un diritto di tutti a camminare nella città e una città camminabile è un diritto di tutti per almeno due ragioni:

La prima ragione deriva dalla nostra visione della camminabilità come un diritto alla città e si basa quindi sulle stesse motivazioni del diritto alla città di Lefebvre. La città è un prodotto della società, si sviluppa grazie alla collettività e i suoi vantaggi e svantaggi devono essere ripartiti nel modo più equilibrato possibile, non a vantaggio di una sola categoria (nel nostro caso gli automobilisti, nel caso di Lefebvre gli industriali). Se per Lefebvre la città era plasmata dalla classe imprenditoriale per noi la città subisce forti influenze a causa dell'automobile in due principali forme:

- la prima deriva dai comportamenti sociali influenzati dall'automobile, parliamo ad esempio della diffusione urbana, dei sobborghi, dell'urbanizzazione delle campagne, del “rurbano” e dello *sprawl*, conformazioni spaziali derivanti da

comportamenti che plasmano la città a misura di automobilista e non di pedone; non possiamo però giustificare una promozione della mobilità lenta con la teoria di Lefebvre, in questo caso la città è *prodotta* dai comportamenti sociali! Lo faremo più avanti con la teoria di Sen (paragrafo 1.4, pag. 22).

- la seconda forma è che la città viene spesso sviluppata, progettata, gestita tenendo conto delle necessità degli automobilisti piuttosto che di quelle dei pedoni. Anche in questo caso secondo due visioni: quella degli amministratori, e quella delle dinamiche imprenditoriali. Possiamo dire che in questi due casi la città è prodotta non dalla società ma da altri poteri che, in un modo o nell'altro, hanno una differente visione della città (valore di scambio): i primi, gestiscono la città secondo una logica di incentivo dell'*uomo-tipo* (auto-munito!), ignorando il diritto alla città delle popolazioni più deboli, che costituiscono la società e hanno perciò il diritto di usare lo spazio comunemente costruito. Questa tendenza deve essere gestita dagli urbanisti in forma diretta, si deve pensare (e lo si comincia a fare) a progettare la città in modo più camminabile (per tutti), questo testo ha lo scopo di provare a cercare qualche soluzione. La seconda visione è invece dovuta alle dinamiche economiche che sfruttano la struttura della città e della società, influenzando i comportamenti sociali, guardando alla città per il suo valore di scambio proprio nell'accezione Lefebvreiana. Un esempio dove ritroviamo la città capitalista – descritta da Harvey – è nei centri commerciali; questi spostano il baricentro urbano nelle zone industriali, nei pressi delle grandi arterie di trasporto, modificando la città che deve adattarsi ai flussi di auto che raggiungono le nuove zone commerciali, impossibili – o per lo meno con sforzi notevoli – da raggiungere a piedi. Questi contribuiscono a costruire una città dei veicoli e in questo caso lo scopo degli urbanisti si può solo limitare alla ricerca di alternative (es. promozione del centro o azioni di incentivo dei centri commerciali naturali). La seconda ragione per cui una città deve essere camminabile trova nuovamente argomentazione nella giustizia come equità¹⁴, ed è che c'è chi – per una ragione o

¹⁴ Non sarà trattata nel testo ma la concezione di giustizia è in linea con il pensiero di Rawls, secondo cui la giustizia è funzione dell'equità e dell'uguaglianza in base alle argomentazioni del contratto sociale.

Per approfondimenti si veda: Rawls, J. (1971). *A theory of justice*. Harvard University Press.

per l'altra – non può avere o usare un'auto e deve poter essere messo in condizioni di muoversi al meglio delle proprie capacità, in condizioni di equità massima.

Rendere una città camminabile non è solo un diritto di chi preferisce la mobilità lenta a quella veloce, ma è un diritto di chi non ha le *capacità* (questa parola non è scelta a caso, vedremo perché) o le disponibilità per muoversi con il mezzo privato. Non avere le capacità sta a significare – in questo caso – non potersi muovere rapidamente, questo accade alle così dette “popolazioni deboli”, ovvero, bambini, adolescenti, anziani, disabili, poveri, immigrati, emarginati, senza-tetto, ecc., ma anche chi non usa l'auto per altri motivi, parliamo di pedoni, ciclisti, corridori, *skaters*, *segway-ers* e “monopattinatori”.

Possiamo quindi giustificare la camminabilità da questo punto di vista, a partire dal concetto di capacità di Sen che tratteremo nel paragrafo successivo¹⁵.

1.3 L'approccio delle capacità di Sen

Possiamo dire che l'approccio delle capacità è un valido sostegno teorico per giustificare gli interventi alla mobilità lenta ma anche un'ottima chiave di volta per intendere la misura formale del benessere individuale.

L'approccio delle capacità può giustificare gli interventi alla mobilità lenta per due ragioni: (1) alcuni comportamenti sociali – figli dell'automobile, come la diffusione urbana – creano situazioni nelle quali l'*abitante-non-tipo*¹⁶ subisce una limitazione delle possibilità di movimento e (2) chi si trova in questa situazione deve poter essere messo in condizioni di pari libertà.

Il *capability approach* trova poi applicazione nel contesto della valutazione e della misurazione del benessere individuale¹⁷, lo discutiamo (anche) poiché, nei

¹⁵ Ci sarebbero anche altri motivi per cui una città deve essere camminabile ma esulano dall'obiettivo di questo testo. Ad esempio parliamo dei vantaggi ambientali, economici e salutari.

¹⁶ La definizione estesa di *abitante-tipo* è riportata nel paragrafo 1.4, con questa espressione si intende l'abitante adulto, maschio, sano, istruito, ricco e auto-munito (Blečić *et al.* 2014b).

¹⁷ Alcuni tra gli esempi applicativi più noti sono contenuti nel rapporto Stiglitz-Sen-Fitoussi e si possono trovare nell'indice HDI costruito dalle Nazioni Unite secondo i principi dell'approccio di Sen.

prossimi capitoli, ragioneremo sul concetto di misura dell'accessibilità e la nostra visione alla misura sarà – in più punti di vista – dipendente dai concetti di Sen.

In questo contesto non serve fare una descrizione puntuale della teoria delle capacità, è utile limitarsi agli aspetti che saranno necessari per giustificare la tesi e per introdurre il tema della misura della camminabilità o, più in generale, della qualità della vita.

La teoria delle capacità¹⁸ sviluppata alla fine degli anni '70 dal premio Nobel per l'economia Amartya K. Sen¹⁹, è considerata come un'alternativa al cosiddetto utilitarismo²⁰; lo citiamo perché sarà utile per un confronto con i più comuni metodi di misura della qualità della vita, in quelli che definiamo approcci di "contabilizzazione". Per usare le parole di Sen «l'approccio utilitaristico inaugurato da Jeremy Bentham [...] ritiene che il modo migliore per stimare il vantaggio di una persona e paragonarlo a quello delle altre sia focalizzarsi sulla felicità o sul piacere individuale (o su qualche altra interpretazione dell'*utilità* individuale). Un'altra impostazione, riscontrabile in molte applicazioni pratiche della teoria economica, calcola il vantaggio di un individuo in base al suo reddito, al suo patrimonio o alle sue risorse. Entrambe queste prospettive, basate rispettivamente sulle risorse e sull'utilità, si contrappongono all'approccio delle capacità, che si fonda invece sulla libertà.» (Sen, 2014, p.241). Sen contesta l'idea di considerare la sommatoria delle utilità ottenute dal fare o dall'essere, come l'unica forma di contabilizzazione del benessere collettivo, che manca di una visione della differente valutazione dell'utilità tra individui (non tutti traggono la stessa utilità da una medesima cosa), della differente distribuzione (l'utilità totale può essere efficiente ma non equa) e della possibilità di accedere all'utilità (non tutti hanno le stesse possibilità di accesso all'utilità). A differenza dei metodi di

¹⁸ *Capability approach* in inglese.

¹⁹ Il primo testo nel quale Sen parla dell'approccio delle capacità è:
Sen A.K. (1979) *Equality of What?, the tanner lecture on human values*, University of Utah Press.

²⁰ Per utilitarismo si intendono le teorie orientate alla massimizzazione del benessere. Una prima versione fu inizialmente proposta dal filosofo Jeremy Bentham nel saggio *An introduction to the principles of moral and legislation*, 1789 e fu successivamente ripresa dal suo studente John Stuart Mill nel 1861 in *Utilitarianism*.

contabilizzazione l'approccio delle capacità misura il vantaggio individuale in ragione delle capacità che la persona ha di fare quelle cose a cui, per un motivo o per l'altro, assegna un valore. Il vantaggio di un individuo in termini di opportunità è da considerarsi inferiore rispetto a quello di un altro se a tale individuo sono date minori capacità – minori opportunità effettive – di realizzare ciò a cui attribuisce valore. La qualità della vita è perciò valutata da Sen non nella quantità di risorse o reddito, ma in relazione alle *capacità* di conseguire *funzionamenti* di valore.

Capacità e funzionamenti (*capabilities* e *functionings*) sono concetti sostanziali dell'approccio. L'insieme dei funzionamenti rappresenta gli stati o le cose che una persona raggiunge o fa, includendo tutto ciò che l'individuo ritiene importante per la vita²¹.

La capacità è invece definita come l'idoneità o l'abilità di prendere parte agli aspetti della vita che l'individuo ritiene importanti. Si riferisce quindi all'insieme dei funzionamenti che la persona può raggiungere o compiere. La capacità può essere intesa anche come potenzialità e opportunità del mondo circostante alle quali l'individuo può liberamente prendere parte o che può usare.

L'approccio delle capacità misura perciò il benessere individuale attraverso una indicazione delle capacità che la persona ha di fare quelle cose a cui assegna un valore. Se ad un individuo sono date minori capacità (cioè minori opportunità effettive) di ottenere (o essere) ciò al quale attribuisce un valore, egli è in una situazione di svantaggio.

1.4 Sul *capability approach*, applicazioni

L'approccio delle capacità è quindi un candidato efficiente per sostenere la tesi di questo testo – la “misura del diritto alla città” – sia da un punto di vista teorico che metodologico. Teorico nel senso che dobbiamo promuovere una città dove gli individui – a prescindere dai *funzionamenti* – abbiano le stesse *capacità* di movimento, siano perciò tutti messi in condizioni di equità e di pari libertà. Metodologico nel senso che non limitiamo la nostra misura alle componenti dello

²¹ Anticipando, guardiamo alla percezione della qualità dello spazio perché può essere ritenuto un modo di guardare ai funzionamenti che gli individui ritengono importanti per la vita (urbana).

spazio (*commodities*²²) a disposizione degli individui, ma ne misuriamo la loro capacità d'uso, e lo facciamo in due modi: indagando le percezioni – che in un certo qual modo possono dirci quali sono i *commodities* e i *functionings* importanti per chi usa la città²³ – e costruendo un metodo di valutazione che non contabilizzi le disposizioni urbane (*commodities*) ma tenga conto della capacità di fruirne.

Ma facciamo un passo indietro, al senso teorico dell'approccio delle capacità. Abbiamo lasciato un punto in sospeso sul perché è utile e giusto promuovere la camminabilità: abbiamo detto che la città è plasmata (anche) dai comportamenti sociali dell'*abitante-tipo*, adulto, maschio, sano, istruito, ricco e auto-munito, che (ad esempio) produce diffusione urbana²⁴ e di conseguenza crea svantaggi per chi non possiede le stesse capacità.

Non possiamo sostenere che questa sia una privazione del diritto alla città, in questo caso la città è comunque prodotta dalla collettività e dai suoi comportamenti (perciò non da forze politiche o economiche), ma chi genera questi processi è una parte della società, quella più forte. Sembra quindi sensato sostenere che la promozione della camminabilità sia un “diritto fondamentale”, poiché, nella misura del benessere collettivo – secondo l'approccio delle capacità – non ci troviamo in una situazione di equità, ovvero c'è chi non ha le capacità per muoversi in libertà.

²² Secondo l'approccio delle capacità sono *commodities* beni e servizi a disposizione degli individui. Si veda: Sen A.K. (1999) *Commodities and Capabilities*, Oxford University Press.

²³ In questo modo è possibile sviluppare una misura che tenga conto delle *capabilities* urbane per gli individui, in un certo modo cerchiamo di rispondere all'idea di Sen di un approccio aperto ed evolutivo (non ritiene opportuno definire una lista predefinita di *capabilities* cui fare riferimento) e di superare (in linea teorica) il limite della teoria di Nussbaum (2011) di indicare quali sono le capacità fondamentali di un individuo (nello spazio). È chiaro però che dal punto di vista pratico il limite diventa quello di costruire un modello che sia esaustivo di tutte le capacità che possono essere ritenute importanti per gli individui e che i modelli di analisi adottati posseggono errori di misura intrinseci.

²⁴ Il riferimento è alla diffusione urbana o *sprawl*, ma anche a tutte quelle situazioni nella quale l'automobilista produce uno svantaggio verso i pedoni, come per la sicurezza, l'inquinamento, la congestione ma anche le pressioni (più o meno dirette) sulla politica per incrementare lo spazio dei veicoli a discapito dei pedoni.

Se la città, come abbiamo detto, è plasmata dalle possibilità che l'auto ci dà – diminuzione della densità, *sprawl*, città diffusa, autostrade urbane, ecc. – ed è modificata dai comportamenti di chi ha le capacità per muoversi secondo gli standard dell'auto, possiamo sostenere che l'utilità che i diversi individui traggono dalla città non è equa. C'è infatti chi non ha le capacità (abbiamo detto bambini, anziani, disabili, ecc.) per godere della città costruita per un altro standard di individuo. Incentivare la camminabilità significa incrementare le capacità di spostamento di quelle popolazioni che si trovano in una condizione di svantaggio dovuta alla loro condizione e/o alla struttura della città.

Abbiamo detto che il concetto di capacità e di funzionamenti è anche alla base della visione della misura di accessibilità in termini metodologici. Ma da un punto di vista spaziale come intendiamo i funzionamenti e le capacità? Come possiamo traslare la teoria di Sen alla pratica urbanistica?

I “funzionamenti urbani” (Talu, 2013) sono costituiti da ciò che l'individuo fa *con* la città e *nella* città²⁵, sono ad esempio l'uso dei servizi pubblici, dei parchi, gli spostamenti da un punto A ad un punto B. Tutti questi aspetti sono facilmente misurabili da un indicatore, in quello che abbiamo definito approccio della contabilizzazione, come il numero di utenti fruitori del servizio, il numero di parchi, il numero di spostamenti da A a B.

Per capacità si intende invece ciò che l'individuo può fare (è libero di fare, ha modo di fare) nello spazio urbano (*con* la città e *nella* città). Ciò che si vuole fare e quindi misurare (anche) questi aspetti, che ci permettono di guardare alla qualità della vita da un punto di vista individuale. Ma prima di vedere come, è necessaria una specificazione; il concetto di capacità comprende infatti due significati: abilità e opportunità (Talu, 2013, p.46). La prima sta ad indicare ciò che l'individuo può fare in funzione delle sue caratteristiche, ad esempio un individuo che non possiede capacità motorie non avrà l'abilità di muoversi a piedi nello spazio. Mentre per opportunità si intende ciò che l'individuo può fare in funzione delle

²⁵ Cioè le attività che svolge grazie alla forma della città e ciò che svolge all'interno della configurazione spaziale della città.

Per approfondimenti si veda: Talu, V. (2013) Qualità della vita urbana e approccio delle capacità, *Archivio di studi urbani e regionali*, 107: 52-73 e Nuvolati, G. (1998) *La qualità della vita delle città, Metodi e risultati delle ricerche comparative*, FrancoAngeli, Milano.

risorse al quale ha accesso, in questo caso significa che un individuo che non ha la capacità di camminare può comunque muoversi nello spazio se ha a disposizione degli strumenti che gli permettono di farlo, ad esempio una carrozzina. Le abilità sono perciò strettamente personali e, nella maggior parte dei casi, non sono influenzabili nella pratica di gestione dello spazio²⁶. Ciò che possiamo fare come urbanisti è intervenire affinché le caratteristiche dello spazio si adattino il più possibile alle abilità (es. orientamento topografico²⁷). Mentre le opportunità si riferiscono, nel nostro caso, alla struttura fisica e spaziale della città e alle sue dotazioni. La pratica urbanistica ha in questo forti responsabilità, può infatti intervenire nel disegno dello spazio fornendo dei *commodities* che siano fruibili da tutti, cioè ampliando le capacità d'uso degli utilizzatori (es. con presenza di parchi giochi [formali e informali], assenza di barriere architettoniche, presenza di pavimentazioni tattili o semafori sonori). Quest'ultimo punto – l'analisi dei *commodities* per misurare le *capabilities* – è ciò di cui tratta questo testo.

²⁶ Un elemento che potrebbe essere considerato abilità è la conoscenza, in questo caso si potrebbe considerare la diffusione delle informazioni come un elemento di gestione delle abilità, come ad esempio l'informazione sul trasporto urbano o sulle disponibilità urbane.

²⁷ Si veda ad esempio: Frau, G. (2015) Natural perceptual wayfinding for urban accessibility of the elderly with early-stage, *Journal of Technology for Architecture and Environment*, 1, 106-113.

Capitolo 2

Come e cosa misurare (dall'accessibilità alla camminabilità)

Abbiamo detto, che da un punto di vista metodologico, il nostro intento è quello di applicare la teoria delle capacità in maniera operativa, vogliamo “misurare il diritto alla città” inteso come diritto alla camminabilità. Ma per ottenere una misura formale della camminabilità abbiamo bisogno di una *proxy*, un elemento che sia misurabile. Quando parliamo di spostamenti nello spazio c'è un concetto che è comunemente condiviso, l'indicatore della capacità di movimento, l'accessibilità. Prima di vedere in pratica come misurare è necessario fare una digressione su cosa sia l'accessibilità, su quale sia la sua differenza con l'accesso e con la camminabilità e infine su come è comunemente intesa e misurata. In questo testo ci limitiamo a esaminare alcune considerazioni di accessibilità intese in un'ottica di equità, che a loro volta richiedono qualche precisazione. Quindi, questo secondo capitolo sintetizza gli elementi chiave della ricerca bibliografica sui concetti di equità, della sua applicazione spaziale (equità spaziale) e delle forme di interpretazione del termine accessibilità, comunemente inteso come misuratore dell'equo accesso allo spazio. Diverse interpretazioni di ogni concetto portano a diversi metodi di misura, che in linea generale possono essere sintetizzati in *place-based* e *people-based*. Ogni metodo si compone di differenti procedure di stima, queste sono elencate per discutere il metodo più appropriato per gli scopi di questo testo²⁸. I prossimi tre paragrafi sono destinati alla descrizione dei concetti di equità (2.1), equità urbana (2.2) e accessibilità (2.3); il rapporto tra accessibilità, accesso e camminabilità è discusso nel paragrafo 2.4 e nel successivo (2.5) si descrivono e analizzano i metodi di misura dell'accessibilità, per concludere con una discussione sulle applicazioni per la nostra misura (2.6).

²⁸ Buona parte di questo capitolo è un estratto dell'articolo pubblicato su Archivio di Studi Urbani e Regionali: Canu, D (2015) Misure di accessibilità nella valutazione dell'equità urbana: una raccolta dei metodi, *ASUR*, 116: 115-139.

2.1 Concetti di Equità

A questo punto è utile trattare il concetto di equità per alcuni motivi. Abbiamo detto che uno dei compiti dell'urbanistica è garantire l'equità (spaziale) tra i cittadini, lo giustifichiamo da un lato con il diritto alla città per tutti, poiché produttori della città, e dall'altro secondo i criteri della giustizia di Sen (e indirettamente di Rawls) accennati in precedenza.

Il concetto di equità come giustizia appartiene ad un'ampia tradizione di filosofia politica²⁹ ed è applicato agli studi urbani del diritto alla città (Harvey, 1973), della giustizia spaziale (Soja, 2010) e della città giusta (Fainstein, 2010; Ischia, 2012).

Ma senza altre precisazioni il vocabolo equità non definisce con chiarezza in che modo le risorse debbano essere distribuite, per questo Nicholls (2001) propone, secondo le tipologie di equità di Lucy (1981), Crompton e Wicks (1988), lo schema in Figura 2.1. L'equità può essere suddivisa in quattro classi e otto sottoclassi.

- L'equità basata sull'uguaglianza può essere operativizzata in due forme. Allocazione di *input* in eguale modo, indipendentemente da condizioni socio-economiche o dalla zona di residenza, secondo un criterio di equa distribuzione che è indipendente dalla situazione iniziale e da quella finale. Allocazione di *output* che, in opposizione, risponde ad un principio di uguaglianza di condizioni, cioè una distribuzione diversa per rendere uguali. Paradossalmente uguaglianza di *input* si traduce raramente in uguaglianza di *output* e viceversa.
- L'approccio di equità compensatoria o basata sui bisogni, può essere riassunto con il concetto "trattare le cose uguali in modo uguale e le disuguali in modo disuguale". Un limite di questo approccio è che la scelta di "chi è svantaggiato" o "chi ha bisogno" rimane soggettiva, questo potrebbe essere argomento di dibattito; solitamente l'indicatore è quello socio-economico come il reddito o una condizione di disabilità.
- Equità in base alla domanda potrebbe presentare questo stesso problema, si basa sul concetto di distribuzione a chi dimostra di avere interesse in un servizio o

²⁹ Non discusse in questo testo, per un'introduzione si rimanda il lettore a Sandel, M. (2010) *Justice, What's the right thing to do?*, Farrar, Straus and Giroux, New York.

struttura. Chi ha bisogno è solitamente definito attraverso osservazione di uso (modello economico) o richieste politiche (modello politico).

- L'ultimo modello di equità è il potere d'influenza delle forze di mercato nella distribuzione dei servizi. L'allocazione delle risorse è in accordo con la quantità di tasse pagate, la disponibilità a pagare o l'alternativa di costo minimo.

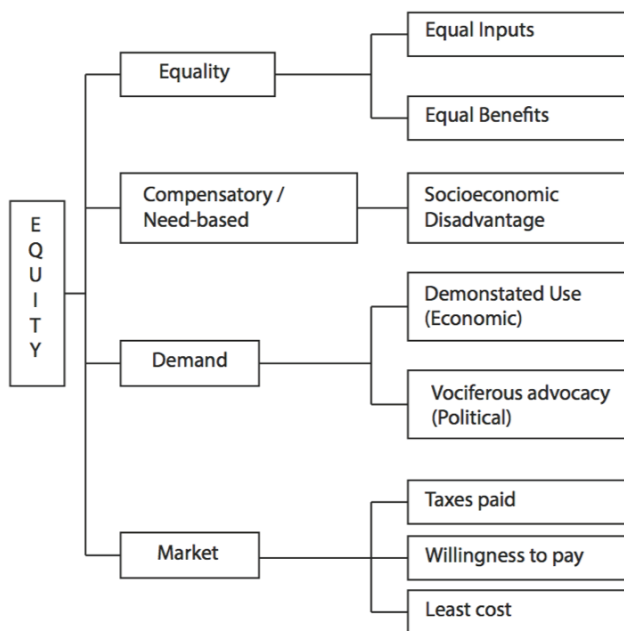


Fig. 2.1 Schema dell'equità. Fonte: Nicholls (2001).

A questo punto ci si potrebbe chiedere: quale tra queste è la definizione preferibile? Scegliere una tra le diverse concezioni di equità varia in funzione di ciò che si sta misurando e secondo i propri valori di giudizio, o per meglio dire, secondo ciò che sarebbe preferibile per la società. Secondo il nostro approccio teorico possiamo certamente sostenere che l'equità compensatoria è intesa come definizione della giusta distribuzione delle risorse, dell'accessibilità e della camminabilità. La teoria dell'approccio delle capacità, come accennato, si fonda infatti sul concetto di equa distribuzione in funzione delle diverse possibilità e capacità. Definire "chi è svantaggiato" rimane un limite dell'approccio a causa della sua ridotta definizione delle *capabilities* fondamentali, che per Sen devono

essere definite caso per caso secondo un approccio aperto ed evolutivo³⁰. Secondo il nostro punto di vista i più svantaggiati nella città sono i cittadini con meno diritti e i non-cittadini³¹. Se per cittadini, come abbiamo già anticipato, consideriamo l'uomo, adulto sano e auto-munito, in accordo con lo schema di Cecchini e Musci (2008) per cittadini con meno diritti intendiamo principalmente gli anziani, i poveri, i senza-tetto, i disabili; per non-cittadini intendiamo gli stranieri (residenti o clandestini), i bambini (che sono non ancora cittadini), i pazzi e i detenuti (non più cittadini). Dal punto di vista della accessibilità (e della mobilità autonoma), possiamo considerare queste categorie come svantaggiate, l'incremento della camminabilità nelle città è – secondo il punto di vista in questa tesi – uno strumento di equità distributiva³².

Questa visione è condivisa anche dalla maggior parte degli studi di riferimento di questo testo ma più variegata sono le scelte di “chi è svantaggiato”. Rispetto agli altri approcci alla definizione di equità nessuno degli autori citati si riferisce ad equità come allocazione in accordo al modello d'influenza delle forze di mercato. Solo alcuni stimano l'accessibilità come misura dell'uguaglianza (sia *input* che *output*), mentre il terzo modello di equità, che distribuisce secondo la domanda, non è citato in nessun caso; si tratta di un indicatore molto difficile da rilevare.

2.2 Definizioni di equità spaziale

Abbiamo appena definito che il nostro interesse è quello di ricercare l'equità spaziale. In generale, il suo obiettivo di ricerca è di accertare che la distribuzione dei servizi pubblici sia equa e correlata ai modelli spaziali socio-economici osservati (Talen e Anselin, 1998). Le ricerche sulle dimensioni spaziali di equità si basano sulle scienze sociali, le loro definizioni e le ambizioni sono perciò varie,

³⁰ Martha Nussbaum prova a definire con più dettaglio questo aspetto, fornisce anche una lista delle dieci *capabilities* fondamentali, in una lista completa e aperta.

³¹ La classificazione preliminare non è esaustiva, mancano ad esempio i non-cittadini per scelta. Per approfondimenti si veda: Cecchini, A., Musci, E. (a cura di) (2008) *Differenti? è indifferente. Capire l'importanza delle differenze culturali e fare in modo che non ci importi*, Edizioni la meridiana, Molfetta.

³² Promuovere la camminabilità porta in realtà anche benefici anche per i già cittadini, in termini di salute, opportunità, qualità della vita, ecc.

così come gli indicatori con i quali i ricercatori hanno tentato di misurare l'obiettivo postulato (Kunzmann, 1998). Equità, in un contesto spaziale, può essere variamente definita e misurata (Hay, 1995) e la determinazione di ciò che è giusto ed equo, varia secondo sistemi di valori di personali (Truelove, 1993). Harnik (2003) propone la definizione di equità spaziale maggiormente condivisa, che riassume chiaramente il concetto come al grado con cui i servizi o gli spazi pubblici sono distribuiti su aree differenti per diversi gruppi economici, etnici e politici, con un'adeguata considerazione alle esigenze dei gruppi speciali (Omer, 2006). Citandolo: "Uno spazio pubblico è accessibile a tutti indipendentemente dalla residenza, dalle abilità fisiche o dalle risorse finanziarie. Questi devono essere localizzati in modo tale che ogni residente sia equamente servito. Inoltre, l'accessibilità non dovrebbe essere basata su un ideale adulto sano, ma piuttosto su un anziano con un bastone, una madre che spinge un passeggino e su un bambino di otto anni che usa una bicicletta" (Harnik, 2003)³³.

Ciò che varia nei differenti approcci è cosa si intende per "tutti" e soprattutto come gli spazi pubblici "devono essere localizzati". Più concretamente, per alcuni, l'equità spaziale è solo la parità di accesso ai servizi pubblici di base misurata in distanze (Kinman, 1999; Ogryczak, 2000, Taleai *et al.*, 2014), come ad esempio l'accessibilità alla scuola, a strutture sanitarie o eventi culturali; è una visione derivante da una concezione dell'equità come uguali *output*.

Secondo altri, l'equità spaziale è più ambiziosa e implica che ci sia una distribuzione uniforme dei servizi in relazione ai bisogni, alla mobilità e ai servizi standard per ciascun residente (Chang e Liao, 2011; Talen e Anselin 1998). La connotazione dell'equità spaziale è che tutti i residenti debbano essere trattati allo stesso modo, indipendentemente dal loro potenziale di mobilità. In questo contesto della pianificazione urbana, equità spaziale significa uguale separazione spaziale o vicinanza a strutture pubbliche tra tutti i residenti.

Altri ancora misurano l'equità spaziale in base al livello di esclusione sociale, definito come il processo che non permette ad un individuo o un gruppo di

³³ "A public space is accessible to everyone regardless of residence, physical abilities or financial resources. They should be sited in such a way that every resident is equitably served. Moreover, accessibility should not be based on an idealized healthy adult but rather on a senior with a cane, a mother pushing a stroller or an eight-year-old riding a bicycle" (Harnik, 2003), traduzione propria.

partecipare alle normali attività della società dove risiedono. L'esclusione sociale non è dovuta alla mancanza di opportunità sociali, ma la mancanza di accesso a tali opportunità³⁴ (Preston e Rajé, 2007).

Questi ultimi due approcci propongono una visione dell'equità come distribuzione compensatoria, nel primo caso secondo le caratteristiche dello spazio, nel successivo secondo le caratteristiche della persona. La nostra misura parte da questi due approcci, ma si sviluppa in maniera composta.

Infatti, l'obiettivo principe di questo lavoro è la costruzione di modelli di valutazione della camminabilità come strumenti per "misurare il diritto alla città e l'equità spaziale", secondo i nostri principi l'equità urbana deve essere misurata sia guardando alle caratteristiche dello spazio (*commodities*) sia dalla loro reale influenza sugli individui in funzione delle loro capacità (*capabilities*) o, più in generale, dei loro desideri, delle loro percezioni. L'approccio alla misura delle caratteristiche dello spazio è quindi la base dei modelli, ma la loro valutazione è funzione delle preferenze dei decisori.

2.3 Definizioni di accessibilità

Serve ora capire in che modo misurare l'equità spaziale, Talen e Anselin (1998) sostengono che gli indici di accessibilità sono stati i più utilizzati, ma seppure l'accessibilità è un concetto di uso frequente, non vi è consenso circa la sua definizione (Lotfi e Koohsari, 2009). È un concetto ben noto nel campo della pianificazione dei trasporti sin dal 1950 quando è stata definita come la facilità di raggiungere destinazioni desiderabili (Hansen, 1959). Più precisamente: accessibilità è comunemente definita come la facilità con cui le attività possono essere raggiunte da un determinato luogo e con un certo sistema di trasporto (Morris *et al.*, 1979; Johnston *et al.*, 2000; Dalvi, 1978; Dalvi e Martin, 1976; Geurs e van Wee, 2004). Si riferisce alla facilità con cui un edificio, un luogo o una struttura sono raggiungibili da persone e/o beni e servizi (Cowan, 2005, p. 2). Handy e Niemeier (1997) hanno suggerito che l'accessibilità è determinata dalla distribuzione spaziale delle potenziali destinazioni e dalla facilità di raggiungere

³⁴ A questo proposito Sen sottolinea che il diritto è una forma di inclusione sociale e la privazione di capacità è una forma di esclusione (Sen, 2009).

ogni destinazione, ma anche dall'attrattività, dalla qualità e dalle caratteristiche delle attività che vi si trovano. Ogni definizione citata comprende inoltre uno o più componenti che influenzano l'accessibilità, ad esempio la posizione di attività o il sistema di trasporto (Vandenbulcke *et al.*, 2009). Secondo una visione più semplicistica e generale, accessibilità è invece intesa come la relativa vicinanza o prossimità di un luogo ad un altro (Chang e Liao, 2011; Tsou, *et al.*, 2005).

Ma accessibilità si riferisce anche alla capacità di un individuo di prendere parte ad una particolare attività o un insieme di attività (Lau e Chiu, 2003); in altre parole, è la capacità delle persone di raggiungere e partecipare alle attività (Garb e Levine, 2002). Da questo punto di vista viene anche intesa come indicatore della possibilità di partecipare alle attività considerate normali per la propria società, come ad esempio l'accesso al lavoro e servizi essenziali (Farrington e Farrington, 2005). È quest'ultima una definizione orientata alle capacità dell'individuo che non può che derivare da una visione di giustizia spaziale come accesso alle opportunità, nonché da una visione della giustizia come distribuzione secondo le necessità.

L'accessibilità può quindi essere definita in vari modi, con approcci soggettivi oppure oggettivi (Lotfi e Koohsari, 2009). Questa distinzione sta alla base dei metodi di misura dell'accessibilità, che come vedremo in seguito si dividono in due principali classi, *place-based* e *people-based*. Secondo un approccio oggettivo, uno spazio pubblico accessibile è quello a cui molte persone diverse possono arrivare, ma anche quello in cui molte persone diverse possono fare molte cose diverse (Bertolini, 1999; Bertolini e Djist, 2003). Si tratta sostanzialmente di una misura delle componenti dello spazio. Mentre secondo un approccio soggettivo, l'accessibilità è un termine relativo a individui differenti (cioè caratterizzati da diverse esigenze, capacità e opportunità), in diversi luoghi e momenti della giornata (Vandenbulcke *et al.*, 2009). In questo contesto accessibilità è la libertà o l'abilità delle persone di realizzare i loro bisogni di base, al fine di sostenere la loro qualità di vita. È definibile anche come la libertà degli individui di decidere o meno di partecipare a diverse attività (Burns, 1979 in Neutens *et al.*, 2010). Recker *et al.* (2001, p. 1) hanno definito l'accessibilità personale come: "La misura delle potenziali abilità degli individui [...] non solo nel raggiungere attività, ma di farlo con sufficiente tempo disponibile per

partecipare a tali attività, in funzione dei vincoli spazio-temporali imposti dai loro impegni quotidiani e dai mezzi di trasporto a disposizione”³⁵.

Secondo questi approcci alla definizione di accessibilità, in letteratura, sono state misurate le capacità di movimento e di accesso equo per diversi soggetti verso vari punti d’interesse. Nella Tabella I sono elencati i principali temi di analisi in corrispondenza con gli articoli citati, che ci saranno utili per definire i metodi di misura dell’accessibilità nei prossimi paragrafi.

Tab. I Corrispondenza tra “temi” valutati e articoli citati.

Accesso e disponibilità a:	Articoli
fornitura di cibo sano	Apparicio <i>et al.</i> , 2007; Cummins e Macintyre, 2002; Pearce <i>et al.</i> , 2008
servizi sanitari	Kinman, 1999; Korn, 1978; Knox, 1978; Luo e Wang, 2003; Rosero-Bixby, 2004; Tanser <i>et al.</i> , 2006; Yiannakoulis <i>et al.</i> , 2013
servizi ricreativi	Crompton e Wicks, 1988; Diez Roux <i>et al.</i> , 2007; Robitaille e Herjean, 2008
scuole secondarie e parchi giochi	Chin e Foong, 2006; Pacione, 1989; Singleton <i>et al.</i> , 2011; Smoyer-tomic <i>et al.</i> , 2004; Talen e Anselin, 1998
parchi pubblici e spazi verdi	Chang e Liao, 2011; Comber <i>et al.</i> , 2008; Coutts, 2008; Hillsdon <i>et al.</i> , 2006; Lindsey <i>et al.</i> , 2001; Nicholls, 2001; Oh e Jeong, 2007; Omer, 2006; Pasaogullari e Doratli, 2004; Talen, 1997; Witten <i>et al.</i> , 2008
servizi di quartiere e urbani	Blečić <i>et al.</i> , 2013; Blečić <i>et al.</i> , 2014a; Lotfi e Koohsari, 2009; Neutens <i>et al.</i> , 2010; Pearce <i>et al.</i> , 2008; Taleai <i>et al.</i> , 2014; Talen, 2003; Tsou <i>et al.</i> , 2005
servizi per immigrati	Truelove, 2000
opportunità lavorative per i lavoratori e lavoratori a basso reddito	Lau e Chiu, 2003; Lau e Chiu, 2004
trasporto pubblico (vs privato)	Benenson <i>et al.</i> , 2011; Bertolini, 1999; Lovetta <i>et al.</i> , 2002; Salonen <i>et al.</i> , 2013
mobilità e uso del suolo	Bertolini, 2003; Dalvi e Martin, 1976; Farrington e Farrington, 2005; Garb e Levine, 2002; Geurs e van Eck, 2001; Geurs e van Wee, 2004; Gutiérrez, 2001; Gutiérrez e Urbano, 1996; ; Iacono <i>et al.</i> , 2010; Linneker e Spence, 1996; Morris <i>et al.</i> , 1979; Vandenbulcke <i>et al.</i> , 2009

2.4 Differenze tra Accessibilità, Accesso, e Camminabilità

Dei riferimenti bibliografici presentati sino ad ora in questo capitolo nessuno fa strettamente riferimento ad un concetto chiave di questo lavoro. Nei testi riportati si parla di accessibilità e accesso, in alcuni casi di accessibilità pedonale, ma senza

³⁵ “A measure of the potential ability of individuals with a household not only to reach activity opportunities, but to do so with sufficient time available for participation in those activities, subject to the spatio-temporal constraints imposed by their daily obligations and transportation supply environment.” (Recker *et al.*, 2001, p. 1), traduzione propria.

mai indicare il concetto di camminabilità (*walkability*) che sta alla base di questo testo. Difatti, quella che segue è una rassegna dei metodi di misura di accesso e accessibilità intesi nell'ottica della giustizia spaziale o dell'equità urbana. Parleremo più avanti delle misure di camminabilità, già i primi due concetti hanno alcune differenze sostanziali che, prima di procedere, è bene analizzare. Sono due definizioni relative all'uso di un qualche servizio e, almeno teoricamente, si riferiscono a due differenti punti di vista, nonché a differenti approcci alla misura. La parola accessibilità deriva dal latino *accessibilitas* e descrive la capacità e le caratteristiche dell'essere accessibile, o la possibilità di facile raggiungimento di qualcosa o qualcuno. Mentre accesso, deriva da *accessus*, da *accedere*, ed indica la possibilità o la facoltà di accedere, indicando l'atto di muoversi entro o verso una destinazione. Perciò, nel nostro caso, il sostantivo accessibilità sta ad indicare le caratteristiche di uno spazio e le sue proprietà nell'essere accessibile, ed in un certo senso nella facilità e possibilità che comporta; mentre l'accesso è in funzione di ciò che un individuo fa, o può fare nella città, da una destinazione ad un'altra. Vedremo quindi che formalmente l'accessibilità indica le caratteristiche di una prospettiva localizzativa, mentre l'accesso indica una prospettiva individuale. Questa distinzione sta alla base delle due principali famiglie di metodi di misura dell'accessibilità, i metodi *place-based* e *people-based*. I metodi formali saranno discussi più in dettaglio nel prossimo paragrafo, per ora ci serve nominarli per introdurre e ragionare sulle differenze del concetto e della misura di camminabilità rispetto all'accesso e all'accessibilità. Infatti nelle misure di camminabilità non si intende una semplice applicazione dell'accessibilità pedonale, è qualcosa di più, che integra molteplici aspetti del percorrere lo spazio a piedi. Camminabilità, che traduciamo dall'inglese *walkability*, è un neologismo composto da *walk+(cap)able+ity*; è quindi costituito dal verbo *walk*, che indica l'azione di muoversi a piedi, di camminare, *(cap)ability*, che indica la capacità, l'abilità, la possibilità di essere o fare qualcosa e, infine, il suffisso *-ity* utilizzato per parole astratte che esprimono stati o condizioni. *Walkability* esprime perciò una capacità di muoversi, a piedi nello spazio, ma anche la capacità dello spazio di essere percorso e attraversato a piedi. La camminabilità include quindi – in qualche modo – sia il concetto di accesso (pedonale) sia quello di accessibilità (pedonale), in una visione composita che sta alla base di questo lavoro (una

misura spaziale di distanze ed elementi qualitativi e quantitativi, calibrata sulle percezioni e caratteristiche individuali).

Se quindi da una parte il concetto di accessibilità si focalizza sull'analisi delle disponibilità urbane, misurando la distanza o la distribuzione degli elementi nello spazio, dall'altra quello di accesso misura come le caratteristiche degli individui deformano lo spazio per loro disponibile (es. prisma spazio temporale) o ne misurano il valore in funzione dell'utilità che ogni individuo ne deriva. La nozione di camminabilità include non solo questi aspetti *place-based* e *people-based*, ma coglie anche altri elementi meno analitici dello spazio; include la relazione tra gli individui e la città, le loro percezioni e la qualità dell'ambiente urbano, la sua capacità di attrarre ed essere partecipe della vita sociale nella città, stimolando gli abitanti nell'uso della città (Blečić *et al.* 2015a; Ewing e Handy, 2009, Porta e Renne, 2005).

La camminabilità si offre in questo modo come una misura della qualità spaziale e delle caratteristiche dello spazio che rispondono (anche) ai requisiti di analisi del diritto alla città e non alla sola capacità di muoversi nella città.

Infatti, la nostra misura non si vuole solo limitare alla stima della possibilità di spostamento nello spazio urbano; secondo i presupposti della teoria di Lefebvre, ciò a cui gli individui hanno diritto è la vita sociale e la vita sociale è caratterizzata dalle sfumature qualitative dello spazio, non solo dalla possibilità di interconnessione tra punti della città, in una mera concezione di mobilità.

Dall'altra parte, in riferimento al *capability approach*, le capacità urbane sono misurate attraverso l'idea della disponibilità delle risorse, dalla percezione che gli individui hanno dello spazio e dalla componente fisica delle distanze (paragrafo 3.2). Possiamo quindi dire che la nostra interpretazione della camminabilità si basa sugli attributi dello spazio, sulle percezioni individuali e sulla distribuzione e disponibilità delle risorse in funzione delle distanze fisiche. Come abbiamo anticipato, la misura formale della camminabilità si fonda sull'analisi delle caratteristiche *place-based* ed integra in se la componente *people-based* attraverso le percezioni soggettive. A questo punto ci interessa approfondire le metodologie di misura dell'accessibilità (parliamo di accessibilità perché è il termine adottato comunemente in letteratura, senza distinzione con accesso) che è adottata come

proxy e componente della misura di camminabilità. Analizzare i metodi ci aiuta a comprendere le scelte metodologiche di base della nostra misura.

Per una lettura più scorrevole, se il lettore ha conoscenze dei metodi di misura dell'accessibilità, è consigliato passare subito al paragrafo 2.6 (pag. 47), ritornando all'analisi dei metodi di misura se necessario.

2.5 Metodi di misura dell'accessibilità

A partire dai metodi di misurazione presenti in letteratura è possibile schematizzare i principali sistemi di stima del valore di accesso ed accessibilità secondo lo schema in Figura 2.2:

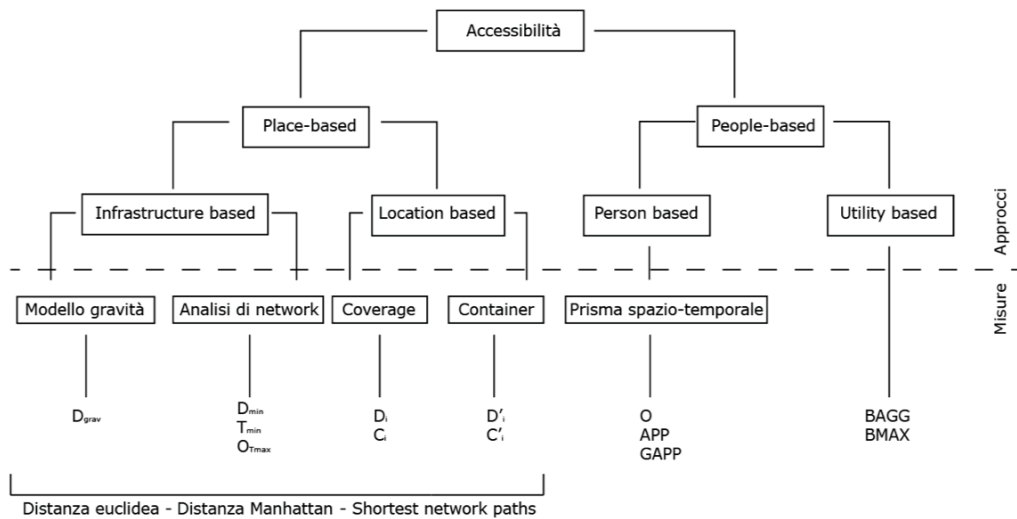


Fig. 2.2 Tassonomia dei modelli di valutazione dell'accessibilità.

Alla ripartizione prevalentemente adottata in letteratura tra misure delle prestazioni dello spazio fisico (*place-based*) e su aspetti del comportamento e dell'utilità (*people-based*) Geurs e van Wee (2004) aggiungono maggior dettaglio fornendo una suddivisione che separa ulteriormente le prestazioni dello spazio in:

- *infrastructure based* (prestazioni delle infrastrutture di trasporto, come “livello di congestione” e “velocità media nella rete stradale”) e;

- *location based* (livello di accessibilità alle attività distribuite spazialmente, come “il numero di posti di lavoro all’interno di un raggio di 30 minuti di viaggio dal punto di origine”);

Mentre distinguono le misure dell’accessibilità a livello individuale in:

- *personal based* (accessibilità al livello individuale come “le attività che un individuo può fare in un determinato arco di tempo”) e;
- *utility based* (benefici che gli individui ottengono dall’accesso alle attività spazialmente distribuite come “la qualità degli spazi o il tipo di servizi presenti nella strada che si percorre”).

A seconda dell’approccio adottato, la misura di accessibilità sarà differente; in termini generali la definizione dell’accessibilità di un individuo i verso un impianto j è una funzione f di: (1) distanza d tra i e j ; o (2) soddisfazione-utilità S che l’individuo deriva dal servizio di j (solitamente stimato utilizzando un vettore di caratteristiche della struttura) (Rosero-Bixby, 2004). Sommariamente la prima definizione corrisponde all’approccio *place-based*, la seconda è invece riassuntiva del modello *people-based*.

Più nello specifico, secondo Koenig (1980) le misure *place-based* sono basate su:

- Opportunità compensate da un impedimento (che è una certa funzione decrescente del costo di viaggio o del tempo per raggiungere queste opportunità), per esempio:

$$A_i = \sum_j O_j f(C_{ij})$$

dove:

A_i = accessibilità dalla zona i al corrispondente tipo di opportunità

O_j = opportunità presenti nella zona j (negozi, luoghi di lavoro...)

C_{ij} = generalmente tempo o costo di viaggio da i a j

$f(C_{ij})$ = funzione di impedimento, generalmente è una funzione esponenziale o funzione di potenza.

- Numero di opportunità che potrebbero essere raggiunte entro un determinato tempo di viaggio x . Questo indicatore può essere considerato una particolare

soluzione alla formula dell'accessibilità A_i con una funzione di impedimento uguale a 1 (per $C_{ij} < x$) o 0 (per $C_{ij} > x$).

La seconda famiglia di metodi di misurazione dell'accessibilità è quella basata sugli approcci del comportamento, detti anche di utilità o *people-based*, che si fondano su due assunti principali:

- le persone associano un'utilità cardinale a ciascuna delle alternative che incontrano (es. con ogni destinazione disponibile, modalità di viaggio, percorso, ecc.) e prendono la scelta associata alla loro massima utilità;
- in quanto non è possibile valutare tutti i fattori che influenzano l'utilità associata ad ogni alternativa per un dato individuo, questa utilità può essere rappresentata come la somma di componenti non casuali (per i fattori prevedibili) e componenti casuali (per i fattori non prevedibili).

L'utilità U_{ij}^t associata all'individuo t , che vive nella zona i e con destinazione j può essere rappresentato dalla formula semplificata:

$$U_{ij}^t = V^t - C_{ij}^t$$

dove:

V^t = utilità lorda nel raggiungere la destinazione j per l'individuo t (variabile casuale)

C_{ij}^t = costo di viaggio generalizzato da i a j per l'individuo t (considerato come prevedibile e non casuale).

Il tempo di viaggio o la matrice dei costi devono essere calcolate appositamente per ogni categoria di individui considerata. Il problema è determinare un'appropriata funzione di probabilità per la variabile casuale V^t e per calcolare il valore atteso dell'utilità massima tra le utilità U_{ij}^t associate a ciascuna delle destinazioni potenziali. Questo valore atteso rappresenta il beneficio medio derivante da un viaggio di un individuo t e tiene conto dell'opportunità di destinazioni, del prezzo del viaggio o del tempo necessario per raggiungerli.

La nostra visione integra entrambe le tecniche, si fonda su una misura delle distanze che include la presenza e la disponibilità di punti d'interesse ma lo fa

integrando nella formula i valori che tengono conto dell'utilità derivata dagli individui.

I sottoparagrafi che seguono (2.5.x), sono una digressione tecnica sui metodi di misura dell'accessibilità *place-based* e *people-based*. Il loro scopo è quello di illustrare le procedure di misura – oltre che per esaustività espositiva – per analizzare gli aspetti di approssimazione modellistica che ogni procedura comporta, così da definire il modello preferibile per il nostro scopo. Anticipando, l'analisi di *network (place-based)* è lo strumento che costituisce la base di misurazione della camminabilità, includendo la componente soggettiva (*people-based*) nella sua calibrazione.

2.5.1 Metodi di misurazione, *Place-Based Measures*

I metodi di valutazione fondati sulla misura delle componenti spaziali sono i più adottati in letteratura, questi si dividono in:

- *Container* (es. numero di aree verdi per quartiere, o parti di esso),
- *Coverage* (es. numero di asili nido in 800 m dal baricentro del quartiere),
- *Network analysis* (analisi del percorso, spesso suddivisa in):
 - Minima distanza (es. la distanza dal centro del quartiere al parco più vicino),
 - Costo di viaggio (es. prezzo del viaggio o tempo di viaggio dal centro del quartiere alla destinazione),
- *Gravity* (es. distanza del percorso tra quartiere e parco tenendo conto del fattore di decadimento).

In riferimento alla classificazione di Geurs e van Wee i primi due metodi sono applicabili alla prospettiva *location based*, gli ultimi due alla *infrastructure based*.

Un problema metodologico che si pone nella ricerca di accessibilità spaziale è la scelta della distanza metrica. In generale ci sono tre tipi di distanze che possono essere usate, ognuna delle quali aumenta il grado di precisione della misura:

- Distanza *Euclidea*: retta.
- *Manhattan distance* (o geometria del taxi): è la distanza lungo due lati di un triangolo rettangolo, la cui base è la distanza Euclidea.
- *Shortest network paths*: percorso di rete più breve.

Di seguito sono proposti alcuni approfondimenti sui modelli; per semplicità espositiva, nelle descrizioni i punti di origine sono indicati come quartieri o case, i punti di destinazione come parchi.

Modello container:

È la definizione della quantità di servizi per abitante, è utile per rapidi calcoli per ottenere un quadro generale dell'adeguatezza del livello di approvvigionamento. Un problema di questo approccio è che assume i benefici solo per i residenti della zona e non considera esternalità spaziali a zone circostanti (Talen e Anselin, 1998). Inoltre non considera la distribuzione spaziale delle opportunità.

Per misurare le prestazioni di servizi urbani in una zona o per un gruppo omogeneo (di individui) si possono identificare due componenti di accessibilità più comuni: la distanza da una residenza ad un servizio e la qualità o la dimensione del servizio (Omer, 2006). Tradizionalmente il grado di fornitura, in accordo con questi due criteri, è determinato da due misure di accessibilità: distanza minima e *container* (per "contenitore" si intende spesso un limite amministrativo o di quartiere secondo fattori socio-economici o geografici).

Rispettivamente misurate con:

$$D_i = \min_j d_{ij}$$

$$C_i = \sum S_j / P_i$$

Dove D_i è la distanza minima tra il centroide (baricentro) del quartiere i (origine) e il parco j (destinazione).

La misura *container* C_i è calcolata dal rapporto tra l'area totale dei parchi S per la popolazione di dimensione P all'interno del perimetro del quartiere i . Questo metodo è spesso criticato perché basato sull'ipotesi che i residenti utilizzino solo i servizi all'interno del proprio quartiere. Inoltre, è possibile che un servizio esterno al quartiere sia più prossimo geograficamente ad un altro localizzato internamente al quartiere.

Modello coverage:

Il secondo modello è basato sulla minima distanza, è una misura dell'area di ricezione che serve il bene pubblico.

Il metodo, nella sua forma più semplice, consiste nell'individuazione del centroide del servizio (es. un parco) che coincide con il centro del cerchio che individua l'area servita dal parco. La dimensione del raggio può variare a seconda delle dimensioni e dalla forza di attrazione del parco (Nicholls, 2001). La distanza minima è calcolata per ogni casa (origine) al parco più vicino (destinazione). L'area di ricezione è misurata dalla quantità di parchi disponibili per ogni casa all'interno dell'area servita dai parchi.

$$D'_i = \min_j d_{ij}$$

$$C'_i = \sum_{d_{ij} < k} E_j; E_j = S_j / \sum_{d_{ij} < k} P_i$$

dove D'_i è la distanza minima tra la casa i ed il parco più vicino j . C'_i , zona di ricezione, è la quantità delle aree dei parchi per ogni persona della casa i all'interno dell'area servita k . Il calcolo di C'_i è composto da tre punti: l'area residenziale è servita dai parchi in accordo con l'area di servizio k (la casa i ha accesso al parco j se $d_{ij} < k$). La densità di potenziali utenti E_j viene calcolata per ogni parco dal rapporto tra la dimensione del parco S_j e la popolazione totale P_i servita dal parco. La quantità di aree parco che coprono una casa è uguale al rapporto cumulativo delle dimensioni di tutti i parchi e delle popolazioni che rientrano nel raggio d'uso dei parchi.

Tre sono i principali limiti di questo modello, che ha d'altro canto il vantaggio di essere un sistema di rapida applicazione: le persone non si muovono nella città in linea retta, questo significa che a seconda della forma della città, dei limiti e ostacoli presenti sul percorso, l'approssimazione può essere più o meno precisa. Il secondo limite è che il modello presuppone che il parco sia accessibile lungo tutto il perimetro. Infine non tiene conto della forma del parco; il modello si adatta ad una forma regolare (quadrata o circolare) ma presenta distorsioni elevate nel caso in cui il parco abbia forme irregolari; inoltre, considerare il baricentro del parco come punto di applicazione distorce la misura all'aumentare della dimensione del

parco. Per sopperire a questi limiti, il modello può essere raffinato; ad esempio, la vicinanza ai servizi può essere misurata combinando il modello *coverage* con l'analisi di *network* (vedi sezione seguente) (Tsou *et al.*, 2005). Oppure, se si applica l'analisi di *network*, il baricentro del parco può essere sostituito con i punti di ingresso (del parco) per ridurre l'errore (Comber *et al.*, 2008).

Network analysis:

Un altro metodo di misurazione, più preciso e maggiormente utilizzato, è basato sulla stima delle distanze lungo le strade tra due punti (solitamente tra residenze e spazi pubblici) attraverso la simulazione del percorso più breve o preferito (es. qualità dello spazio). Questo modello è chiamato *Network Analysis*. Sono principalmente due le applicazioni del modello, la metodologia *network distance* (D_{min}), che misura la lunghezza del percorso di rete più breve lungo le infrastrutture, ed il metodo *travel time* (T_{min}), che considera anche la velocità raggiungibile dal mezzo di trasporto a disposizione dell'individuo. Formalmente:

$$D_{min} = \min_j d_{ij}$$

$$T_{min} = \min_j d_{ij} \wedge \min_j t_{ij} \quad \text{oppure} \quad T_{min} = \min_j t_{ij}$$

Dove d_{ij} è la distanza tra l'origine i e la destinazione j , t_{ij} è il tempo tra l'origine i e la destinazione j .

Una terza misura di accessibilità è O_{Tmax} esprime il numero di opportunità rispetto ad un definito tempo di viaggio massimo.

$$O_{Tmax} = \sum_j P(t_{ij})$$

dove:

$$P(t_{ij}) = 1, \text{ se } t_{ij} < t_{max}; \quad 0, \text{ altrimenti}$$

O_{Tmax} considera non solo l'alternativa più vicina, ma anche tutte le altre opportunità che possono essere raggiunti entro il tempo di viaggio specificato. L'analisi di *network* è un modello che permette di rappresentare abbastanza dettagliatamente la realtà, seppur con qualche difficoltà e approssimazione (si veda paragrafo 2.5.3). La sua applicazione ha però lo svantaggio di richiede competenze e *software* di calcolo specifici.

Modello a gravitazione:

L'ultimo modello D_{grav} è basato sul principio di gravita, è calcolato secondo la formula:

$$D_{grav} = \sum_j a_j \exp[-\lambda_m \min(t_{ij})]$$

dove a_j è l'attrattività dell'attività in j , λ_m è un parametro della distanza di decadimento del mezzo di trasporto m .

In confronto con O_{Tmax} , D_{grav} incorpora due componenti aggiuntivi: l'attrattiva del luogo j e il costo di separazione fisica dalla posizione di riferimento, espressa da una funzione di decadimento per ogni mezzo di trasporto. Essendo il modello a gravità una variante dell'analisi di *network*, si può sostenere che ne condivide pregi e difetti.

2.5.2 Metodi di misurazione, *People-Based Measures*

Le misure *people-based* comprendono misurazioni di accessibilità spaziotemporali, come il tempo di viaggio disponibile e dipendono dalle caratteristiche del sistema di trasporto o della persona.

Il modello più comune è quello del prisma spaziotemporale (*PST*) che definisce tutti i punti spaziotemporali che un individuo può raggiungere in base ad un determinato budget di tempo. Formalmente:

$$PST = \{(j, t) | (t_i + t_{pij} \leq t + \hat{T} \leq t_{i+1} - t_{jpi+1}), e(t_j^0 \leq t + \hat{T} \leq t_j^e)\}$$

dove t è il tempo di attività, t_{pij} è il tempo di viaggio dalla posizione p_i al luogo di attività j ; t_{pi+1} è il tempo di viaggio dall'attività j alla posizione successiva p_{i+1} ; \hat{T} è la durata minima dell'attività; t^o_j è l'ora di apertura dell'attività j , e t^c_j è l'ora di chiusura dell'attività.

In altre parole PST raccoglie tutte le attività j dove un individuo x può svolgere un'attività di durata \hat{T} che rientra nelle ore di apertura della struttura situata in j . La proiezione del PST è definita area potenziale di percorso (APP). Coppie successive di attività fisse possono essere sovrapposte per formare l'area potenziale giornaliera ($GAPP$). L'insieme delle opportunità disponibili O è dato da:

$$O = \{j | (j, t) \in PST\}$$

Neutens *et al.* (2010) propongono due misure derivanti da questo approccio: la prima stima il numero di opportunità nel set O , come la somma del numero di opportunità disponibili (con o senza parametro di decadimento); la seconda è una misura del tempo massimo che un individuo può spendere per raggiungere un'opportunità durante il giorno in funzione del tempo disponibile.

Le misure incentrate sul concetto di utilità sono particolarmente rare, principalmente per la difficile definizione di cosa si intende per utilità. Ad esempio Neutens *et al.* (2010) suggeriscono due metodi di valutazione che misurano rispettivamente l'utilità aggregata delle opportunità nello spazio $GAPP$ e la massima utilità delle opportunità nel medesimo spazio. La prima misura (utilità aggregata), riferendosi a una misura di benefici associati, sarà più grande se il set di localizzazione scelta entro il $GAPP$ contiene più alternative. La seconda (massima utilità) massimizza (invece di aggregare) i vantaggi che un individuo può potenzialmente raggiungere dalle opportunità in O . I modelli *people-based* permettono un'applicazione che include le componenti di preferenza individuale, che rendono differenti quelle scelte che sarebbero altrimenti obiettivamente uguali (a differenza dei modelli *place-based* che si limitano alla realtà oggettiva). Il principale limite è che, trattandosi di preferenze soggettive, è impossibile ottenere una stima priva di errori.

2.5.3 Alcuni approfondimenti sul Modello *Network Analysis*

Il modello dell'analisi di *network* è sicuramente il più adottato, di questo esistono diversi metodi applicativi, tanto più accurati quante più sono le informazioni disponibili o stimate. L'applicazione del modello varia a seconda del mezzo di trasporto che si vuole analizzare, solitamente a piedi, in bici, in auto o con il trasporto pubblico. La simulazione del tempo o della distanza percorsa a piedi può raggiungere un elevato livello di dettaglio, alcuni esempi in letteratura considerano la presenza di strisce pedonali, sovrappassi e sottopassaggi che aumentano il tempo di percorrenza (vedi Oh e Jeong, 2007) o di elementi di arredo urbano e pratiche d'uso che accrescono il valore di qualità dello spazio (Blečić *et al.*, 2013; 2014a). Comunemente, per qualsiasi mezzo di trasporto, il percorso è stimato attraverso algoritmi *shortest path*, che minimizzano il tempo o la distanza di viaggio, oppure massimizzano l'utilità.

Il tempo di viaggio con auto privata è misurato in base alla velocità media o massima di ogni segmento. Il modello può essere perfezionato considerando la congestione, il tempo per la ricerca del parcheggio e l'eventuale tragitto a piedi. Yiannakoulis *et al.* (2013) hanno dimostrato che il tempo di viaggio varia radicalmente il risultato rispetto alla misura delle distanze quando include questi fattori nel modello a gravità.

I tempi di viaggio con mezzi di trasporto pubblico sono soggetti alle regole di frequenza oraria e giornaliera. Tipicamente i tempi sono calcolati semplificando la velocità di viaggio lungo il tragitto ad una velocità media e al tempo medio di trasferimento da una linea all'altra; recentemente si è però potuto misurare con maggior dettaglio grazie a servizi elettronici³⁶.

Anche la definizione dei punti di origine e destinazione influenza il risultato: per esempio, si può semplificare la misura considerando il baricentro di un quartiere, o raffinare la stima adottando come punti di origine tutte le case. Il metodo che simula con maggior dettaglio la realtà è l'approccio *door-to-door*. Può considerare tutti o solo alcuni dei fattori appena descritti, è schematizzato in Figura 2.3. Via

³⁶ Per approfondimenti si veda: Salonen, M., Toivonen, T. (2013) Modelling travel time in urban networks: comparable measures for private car and public transport, *Journal of Transport Geography*, 31: 143-153.

auto il modello include: (1) camminata dal punto di origine al punto nel quale è parcheggiato il veicolo (2) guida sino al punto di destinazione (o parcheggio) (3) ricerca di uno spazio per parcheggiare vicino alla destinazione, e (4) camminata dal punto di parcheggio al punto di destinazione. Via mezzo di trasporto pubblico include: (1) camminata dal punto di origine alla fermata (“tempo di accesso”) (2) attesa del veicolo e/o tempo di attesa prima della partenza (3) tempo di viaggio (4) camminata dalla fermata al punto di arrivo (“tempo di uscita”).

Nel caso in cui il tragitto con trasporto pubblico richieda l’uso di più mezzi, il modello può implementare il tempo di camminata da una fermata all’altra ed il tempo di attesa e viaggio del successivo veicolo.

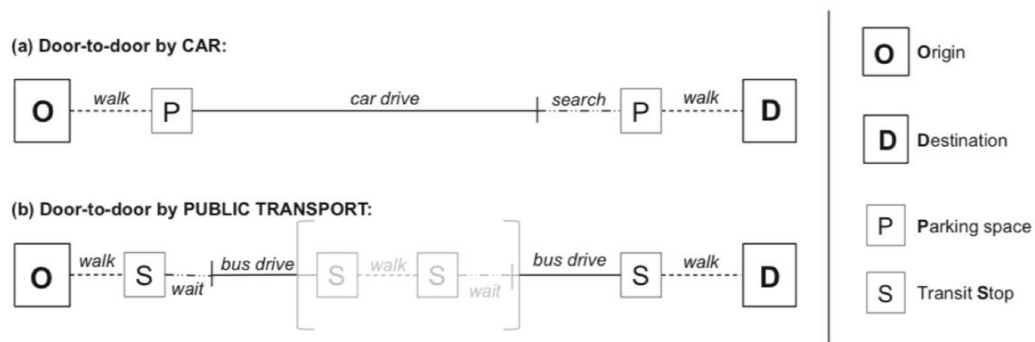


Fig. 2.3 Schema approccio door-to-door. Fonte: Salonen e Toivonen (2013).

A partire dalle procedure di misurazione si possono ottenere due principali risultati: la prima è una misura di accesso a determinati punti di interesse, la seconda è una misura di accessibilità generale per tutta la città.

Nel primo caso si individuano i punti di interesse (es. parchi pubblici) come destinazioni (D), mentre i punti di origine (O) sono solitamente semplificati su una griglia di punti regolari (spesso 250x250m) che copre la città (es. Salonen e Toivonen, 2013; Li *et al.*, 2011); in alcuni studi i punti di origine sono stati indicati per tutte le case della città, il limite a questo approccio è di tipo computazionale (es. Omer, 2006).

Nel secondo caso, misure di accessibilità generale, i punti di destinazione coincidono con quelli di origine. Si ottiene così un valore di accessibilità da e per ogni punto o area. La scelta dei sistemi di calcolo (e quindi di approssimazione) è fortemente dipendente dai dati disponibili e dagli strumenti per misurarli.

2.6 Discussione

La raccolta bibliografica che compone questo capitolo ci è utile per definire una misura di camminabilità che rispetti i presupposti teorici discussi nel primo capitolo.

Abbiamo visto che intendere i concetti teorici secondo differenti punti di vista porta a misure e metodi di stima differenti. La Figura 2.4 mostra i possibili percorsi concettuali che, a partire dalla definizione di equità, arrivano ai due approcci alla misura, *place-based* e *people-based*. Per fare un esempio, considerare l'equità come distribuzione di necessità porterà ad una concezione dell'equità spaziale come distribuzione di pari opportunità o di pari accesso alle opportunità; la prima è una prospettiva di misura dello spazio, la seconda delle caratteristiche della persona. Intendere invece l'equità come uguali *output* significa valutare l'equità spaziale come pari accesso, in una visione *place-based*.

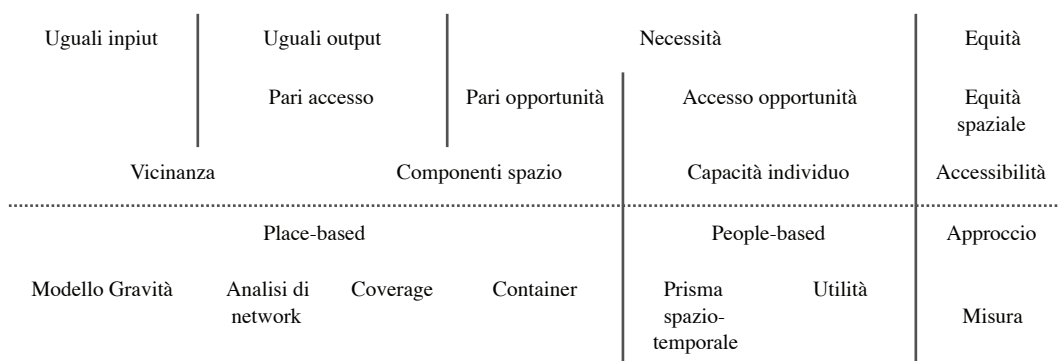


Fig. 2.4 Percorso dalla definizione di equità al metodo di misura dell'accessibilità.

Nel nostro caso, rispetto al tema della giustizia, abbiamo già commentato che – in linea con l'approccio Seniano – la giustizia distributiva rispecchia l'idea di una città giusta intesa come lo spazio di vita di tutti, comprese le popolazioni più deboli (pedoni e non-cittadini). L'equità spaziale può essere perciò intesa come la distribuzione delle risorse che permetta a tutti gli individui di adoperare lo spazio in egual modo, in funzione delle proprie capacità. In termini di camminabilità c'è equità spaziale quando tutti hanno le migliori possibilità di movimento e d'uso dello spazio. Queste visioni non ci consentono dunque di rimanere negli schemi di interpretazione dell'equità spaziale, ma ci portano su un'altra dimensione, una

visione composita che include sia le proprietà dello spazio sia capacità, caratteristiche, preferenze e percezioni dei diversi individui. È questo uno degli elementi che differenziano l'approccio di questo lavoro dai comuni metodi di misura dell'equità urbana, della giustizia spaziale, del diritto alla città e dagli strumenti generalmente adottati per misurare l'accessibilità. La camminabilità è quindi definita sia dalla distribuzione delle destinazioni, dalla loro qualità e dall'attrattività, ma anche dalle capacità di raggiungerle e dall'importanza che ogni categoria di individuo dà agli elementi dello spazio. Secondo questa visione il metodo di misura deve essere ibrido tra *place-based* e *people-based*; così è, la misura è formalmente *place-based*, poiché si basa sulla stima delle caratteristiche dello spazio, ma lo fa secondo una visione di importanza individuale (percezione) degli elementi dello spazio (*people-based*). Il percorso logico seguito sino a questo punto sta alla base di due elementi: (1) il metodo di misura della camminabilità adottato nel modello e (2) la sua calibrazione. Il capitolo 4 sarà interamente destinato alla calibrazione e validazione del modello – il nocciolo di questo lavoro – che si basa sullo studio dell'interazione tra le caratteristiche dello spazio (componente *place-based*) e la percezione degli individui (componente *people-based*). Serve definire (1) quali aspetti dello spazio sono importanti nella valutazione e (2) ricercare un metodo di predizione della qualità – inferito da esempi di assegnazione – che classifichi segmenti stradali, attraversamenti, percorsi, quartieri o città, a partire dalle caratteristiche dello spazio. Il metodo di misura su cui si fonda il modello coincide con gli aspetti di una misura delle caratteristiche urbane in funzione dell'utilità che un individuo deriva dall'uso dello spazio. Il modello è basato sull'analisi di *network* che, come descritto nella parte tecnico-teorica, è il metodo che meglio approssima la realtà, simulando con sufficiente dettaglio il comportamento di un pedone. Ma prima di procedere con l'analisi del metodo di misura e della sua calibrazione è importante guardare allo stato dell'arte dei criteri di valutazione della camminabilità, ad alcuni studi simili e alcuni modelli analoghi, che si fondano sulla misura della camminabilità e non della sola accessibilità o accesso.

Capitolo 3

La misura della camminabilità

Questo capitolo ha lo scopo di riassumere ed analizzare la letteratura di riferimento rispetto alle misure di camminabilità e rispetto alle indagini empiriche di analisi della relazione tra la percezione soggettiva e le caratteristiche dello spazio, che stanno alla base della validazione della misura di camminabilità.

In particolare il capitolo si compone di tre parti:

- la prima (3.1), di analisi degli strumenti di camminabilità, analizza pregi e limiti dei modelli proposti in letteratura, con particolare attenzione a tre elementi principali di ogni modello: la scala spaziale di analisi e di descrizione³⁷, il metodo di aggregazione degli attributi dello spazio che compongono il valore di qualità e la procedura di misura dell'accessibilità in riferimento agli approcci *place-based*.
- la seconda parte (3.2) è di presentazione della misura di camminabilità dalla quale partiamo per definire gli esperimenti del capitolo successivo (il 4°), che servono per validare il modello.
- la terza parte (3.3) analizza la letteratura di riferimento degli esperimenti che indagano percezioni e caratteristiche dello spazio, di questi ci interessano tre aspetti principali: quali sono le variabili dipendenti, cioè qual è la domanda che stima la percezione; quali sono le variabili indipendenti cioè quali sono le caratteristiche dello spazio analizzate e il loro metodi di campionamento; quale è il metodo di indagine statistica adottato (che tipo di regressione) e quali sono i risultati.

3.1 Gli elementi della misura di camminabilità

Al di fuori delle misure di giustizia spaziale, equità urbana e diritto alla città, la camminabilità è misurata per diversi scopi e la sua stima diventa sempre più un

³⁷ Ciò che ci interessa è una scala spaziale molto dettagliata, per questo i parametri (attributi) dei modelli dovranno essere descrittivi delle caratteristiche del disegno urbano. A questo punto però non è importante trattare questo aspetto, che sarà discusso nella terza parte (paragrafo 3.3.1).

tema *mainstream*. Gli articoli sulla *walkability*, “*walkability + measure*” e “*walkability + evaluation*” sono aumentati in circa tre lustri³⁸ del 1.500%³⁹ e sono studiati nei più svariati campi di ricerca, dalla medicina individuale alla sanità pubblica, dalle scienze sociali a quelle ambientali, dall’ingegneria della mobilità e trasporti alla modellizzazione dei comportamenti di viaggio, dalle scienze computazionali applicate all’urbano alle scienze umanistiche. In questi campi la camminabilità è analizzata da più punti di vista, possiamo distinguerli in tre principali gruppi che si sovrappongono e si intrecciano, al quale questo lavoro appartiene in modo più o meno diretto.

Il primo è di studio sulla relazione tra i comportamenti nello spazio urbano e le caratteristiche fisiche (qualitative e quantitative) dell’ambiente pedonale, per indagare e capire quali sono gli elementi dello spazio che influenzano i comportamenti e, perciò, su quali aspetti concentrare la progettazione dello spazio pedonale. Il secondo – a partire da una concezione che possiamo associare all’approccio della capacità – affronta l’importanza delle differenze tra gli individui, i valori e giudizi personali, le preferenze e i comportamenti dei pedoni, la loro utilità o le percezioni dello spazio. Questi studi sono orientati ad una misura dei fattori che gli individui interpretano come importanti nella loro esperienza di camminabilità pedonale (Handy e Niemeier, 1997); si concentrano spesso su particolari popolazioni di non-cittadini, come – tra i più studiati – bambini (Bejleri *et al.* 2010; Buck *et al.* 2011; Gallimore, 2011; Nasar, 2015; Evers, 2014) adolescenti (Yan, 2010), anziani (Garin *et al.* 2014; Towne *et al.* 2016; Van Holle *et al.* 2015) e ciclisti (Nielsen *et al.* 2013; Koh e Wong, 2013). Il terzo include gli studi di camminabilità in termini più generali di capacità individuali e qualità della vita urbana. In questo senso la camminabilità è uno strumento per valutare le capacità e la qualità della vita degli individui, riferendosi al modo in cui possono raggiungere a piedi servizi, attività e opportunità.

Questo studio include i tre campi attraverso una misura della camminabilità che ha l’obiettivo finale di valutare la qualità della vita in termini di diritto alla città e

³⁸ Per l’esattezza si fa riferimento a tre lustri e un anno: 2000/2005, 2006/2010, 2011/2015 e 2016.

³⁹ Valori della somma dei risultati su *Scopus*, *Sciencedirect* e *Google Scholar* per le ricerche con le tre parole chiave.

quindi di uso dei servizi, attività e opportunità della città e della sua vita sociale. Lo fa da una prospettiva di analisi incentrata sulle caratteristiche degli individui, in particolare della loro percezione di qualità dello spazio e dei loro caratteri. Infine, attraverso la relazione di percezione dello spazio e le sue proprietà, individua quali elementi influenzano la progettazione dello spazio e i parametri di un modello di stima della camminabilità.

Oltre agli obiettivi della misura, nella definizione di un modello di camminabilità ci sono tre elementi chiave che andrebbero discussi: la scala spaziale (di *input* o *output*), gli attributi del modello e il metodo di aggregazione dei dati (che comprende anche la procedura di misura dell'accessibilità). Di questi tre aspetti lasciamo da parte gli attributi che costituiscono la base della misura, ne parleremo più approfonditamente nei paragrafi successivi (a partire dal paragrafo 3.3.1), per ora è utile analizzare la scala e i modelli di aggregazione, per esaminare alcuni strumenti disponibili in letteratura sui quali ragionare per definire la nostra misura.

Per scala spaziale si intende il livello di dettaglio degli elementi (i criteri) della valutazione, cioè quanto è dettagliata la misura della camminabilità in termini di inclusione degli elementi spaziali o delle caratteristiche dello spazio analizzato. Ad esempio una *macro*-scala include misure macroscopiche della forma dello spazio ad un livello di blocchi censuari o quartieri, ad esempio analizza il grado di interconnessione urbana, la densità di incroci, la densità abitativa; mentre una *micro*-scala include più dettagli dello spazio, analizzando in maniera puntuale gli elementi nei segmenti e nodi, sino ad arrivare alla valutazione qualitativa degli elementi di arredo urbano, come la larghezza dei marciapiedi, la qualità della pavimentazione, la densità di alberi. La scala di *output* è invece relativa al dettaglio con cui sono visualizzati i risultati, possiamo individuare quattro principali livelli: livello del quartiere, blocchi censuari, livello stradale o sub-stradale (nodi, segmenti, pixel). Le due scale sono strettamente legate, ci serve porre l'accento sul loro livello perché, per il nostro scopo, è utile un modello che adotti una *micro*-scala e quindi *output* con un alto livello di dettaglio, stradale o sub-stradale.

Per modelli di aggregazione si intende il modo in cui gli attributi della valutazione sono aggregati per generare un unico indice di camminabilità. Possiamo distinguere i modelli in compensativi e non-compensativi (ne parleremo più approfonditamente in 4.1.3, ma in breve), i primi - che includono modelli lineari e non lineari - compensano cattivi valori con ottimi valori dei criteri in esame, mentre quelli non-compensativi permettono di tenere conto di questi aspetti attraverso l'implementazione di regole di assegnazione, soglie di maggioranza e percorsi di classificazione. Al di fuori dei nostri test non troviamo in letteratura modelli non-compensativi, ma vedremo che (paragrafo 4.1.3), per i nostri scopi, questa è una proprietà desiderabile che migliora i risultati dei modelli. Infine, un ultimo elemento di aggregazione dei dati è l'analisi dei percorsi. Quando le misure non si limitano ad indicare l'appartenenza di un punto dello spazio ad una categoria (es. un segmento stradale) si adottano gli strumenti di misura dell'accessibilità (*place-based*) visti nei paragrafi precedenti. Quelli che ci interessa analizzare sono basati sull'analisi di *network*, con o senza fattori di decadimento e quelli definiti areali che fanno riferimento ai metodi *container* e *coverage*.

3.1.1 Alcuni strumenti di misura

Perciò, prima di presentare il nostro approccio alla misura è utile spendere le prossime quattro pagine per introdurre alcuni esempi dei più significativi strumenti di misura della camminabilità, cercando di limitare la descrizione ai soli aspetti interessanti per questo testo ed evidenziando pregi, limiti e carenze secondo la linea teorica seguita sino ad ora. Una sintesi è proposta in Tabella II.

Partiamo dal metodo più adottato, divenuto un *mainstream*, *walkscore*; questo strumento si propone come un metodo di misura della camminabilità per trovare il miglior quartiere dove abitare e per la stima dei valori immobiliari. Il suo algoritmo di calcolo stima un punteggio di camminabilità di aree o percorsi in base alla distanza da tredici categorie di servizi⁴⁰, i punteggi sono assegnati in base alla lontananza di ogni categoria di opportunità urbane all'interno di 30

⁴⁰ *Grocery stores, coffee shops, restaurants, bars, movie theaters, schools, parks, libraries, bookstores, fitness centers, drugstores, hardware stores, clothing/music stores.*

minuti a piedi (da 5 a 30 minuti i punteggi sono corretti con una funzione di decadimento). *Walkscore* include misure fisiche della strada come la densità di intersezioni, lunghezza degli isolati e densità di popolazione, il punteggio finale (normalizzato da 0 a 100) è ottenuto sommando i valori di ogni categoria. Questo modello soffre di tre limiti: il primo è che non è un metodo compensativo (che abbiamo detto essere un pregio), il secondo è che per misurare la camminabilità si riferisce al conteggio dei servizi ma esclude altri aspetti come componenti critici, qualità dell'ambiente e percezioni. Tuttavia Carr *et al.* (2010) hanno dimostrato che esiste una correlazione tra il punteggio di camminabilità del modello e alcune misure oggettive del contesto dell'attività fisica nello spazio (*physical activity environment*), di connessione (*street connectivity*), densità abitativa e accesso al trasporto pubblico. I risultati appaiono però anche correlati (positivamente) con il numero di crimini denunciati (*reported crimes*), illustrando un'inabilità a servire come misura assoluta della camminabilità; gli autori consigliano infatti di applicare *walkscore* come una *proxy* per stimare la densità di accessi. Ciò nonostante il risultato è positivamente correlato con la percezione fisica dell'ambiente, anche se gli autori rimangono prudenti su questo aspetto poiché il rischio di correlazione spuria⁴¹ o indiretta⁴² è alto; i risultati possono infatti essere causa di un'elevata presenza di elementi che rendono le aree più percorribili, come ottima illuminazione e ampi marciapiedi. Il terzo punto è che il metodo ignora totalmente la *micro*-scala, cioè non tiene conto degli aspetti di progettazione dello spazio e arredo urbano che influenzano la qualità della camminabilità, come può essere la larghezza di un marciapiede, la presenza di ostacoli nel percorso, la densità di traffico o la qualità visiva dello spazio.

⁴¹ Per correlazione spuria si intende un errore nell'interpretazione dei risultati di una regressione, si verifica quando due variabili non hanno nessi causali o quando le osservazioni non sono indipendenti e identicamente distribuite. Un esempio è il numero di nascite dei bambini e il numero di cicogne che in alcune zone appare direttamente correlato. La reale relazione sta nel fatto che nelle zone rurali le famiglie fanno più figli e nelle medesime zone nidificano più cicogne.

⁴² Anche la correlazione indiretta è un errore di interpretazione dei risultati di una regressione, si verifica quando due variabili risultano correlate da un rapporto di causa-effetto ma in realtà sono indirettamente influenzate da una terza variabile che le collega. Un esempio è il valore del QI tra individui bianchi e neri. Pare esserci una correlazione diretta tra la razza e il valore di QI, la realtà è che il valore di QI è influenzato dal livello di istruzione; la razza è responsabile della bassa istruzione e la bassa istruzione provoca il basso punteggio QI.

Un secondo interessante strumento è quello proposto dal *Future Cities Laboratory*, il loro approccio va oltre la sola distanza dei servizi disponibili ed include nella misura della camminabilità le caratteristiche fisiche dello spazio pedonale (come larghezza del marciapiede, presenza di ombra, inquinamento acustico, separazione dal traffico, ecc.). L'indice di camminabilità è definito dalla funzione ottenuta adottando un *MNL (multi nominal logit) route choice model* che determina il grado di utilità di ogni segmento in funzione dell'utilità stimata attraverso un processo di modellizzazione delle preferenze (*choice modeling*), per includere solo gli elementi importanti per la valutazione di qualità dello spazio. Ciò che manca in questa analisi è, al contrario delle altre, lo studio delle disponibilità di servizi che compongono lo spazio urbano⁴³ e – per rifarci al nostro approccio teorico – alla capacità di raggiungerli ed usufruirne. Inoltre, ciò che importa in un percorso non è solo l'arco stradale, ma anche gli attraversamenti influenzano la scelta dei percorsi, questo modello non tiene conto di tale fattore.

Walkshed è un altro strumento di valutazione della camminabilità che si differenzia da *Walkscore* per la possibilità da parte dell'utente di personalizzare il peso delle variabili che ritiene importanti per la sua idea di camminabilità. Anche in questo caso gli indicatori si limitano ad una analisi dei servizi disponibili senza tenere conto delle caratteristiche fisiche dello spazio e l'algoritmo di calcolo è basato sul metodo compensativo di somma pesata degli attributi.

Walkability level of service è uno strumento di classificazione che permette di aggregare le caratteristiche fisiche dello spazio per ottenere una valutazione del livello dei servizi⁴⁴ attraverso la media dei valori di ogni attributo. Il limite di questo strumento – secondo la nostra prospettiva – è che il modello di valutazione risulta troppo debole e approssimativo, soffrendo in alcune caratteristiche che sono per noi desiderabili, come la non-compensabilità e la validazione dei risultati. Il suo vantaggio sta nel fatto che la sua semplicità lo rende uno strumento applicabile da tutti, anche dai non esperti.

⁴³ La disponibilità dei servizi è inclusa nella funzione di calcolo del modello MNL come densità di servizi. In questo modo l'informazione diventa un attributo della qualità dello spazio e non un elemento di analisi della disponibilità di servizi.

⁴⁴ Per approfondimenti sul concetto del livello dei servizi (*level of services*) si veda il paragrafo 3.3.

Walkscope è invece un interessante strumento (partecipativo ed inclusivo) di resoconto e segnalazione (in una parola *report*) della qualità dei marciapiedi e delle intersezioni. Lo strumento si limita a riportare su mappa le informazioni inviate dagli utenti, fornendo la vista di ogni attributo senza effettuare nessuna operazione di aggregazione per ottenere un valore unico di qualità (il funzionamento è perciò simile a quello di un Web-GIS). Riportare questo esempio è interessante perché, seppure limitato rispetto agli altri, propone un approccio alla valutazione della camminabilità ad un'ampia scala di dettaglio, considerando, oltre la valutazione percepita di qualità per ogni arco stradale o intersezione, informazioni come ad esempio il tipo di marciapiede, la presenza di ostacoli, il volume di traffico, la presenza di panchine e alberi, il numero di corsie da attraversare, la presenza e lo stato di funzionamento dei semafori.

Tab. II Sintesi delle caratteristiche dei modelli di misura della camminabilità.

Modello	Scala input	Scala output	Aggregazione	Tipo analisi	Calibrazione
Walkscore	macro	Stradale	compensativo non-lineare	network + decadimento	comportamenti osservati
Walkshed	macro	pixel	albero di decisione	aree	preferenze utente
Future Cities Laboratory index	micro	sub-stradale	compensativo lineare	network + decadimento	preferenze dichiarate
Walkability level of service	macro	stradale	compensativo lineare	-	preferenze dichiarate
Walkscope	micro	stradale	— —	-	-
Walkonomics	macro	stradale	?	network	?
PEDS Audit	micro	punti	compensativo non-lineare	-	preferenze dichiarate
Ratemystreet	macro	punti	-	-	preferenze utente
ESRI index	macro	blocchi censuari	compensativo non-lineare	aree	-
News index	macro	blocchi censuari	compensativo non-lineare	aree	comportamenti osservati

La Tabella II riassume i più noti strumenti di misura e riporta gli elementi della scala di *input* e *output*, il metodo di aggregazione, il tipo di analisi spaziale e il metodo di calibrazione. A parte i requisiti sulle scale di *input* e *output*, il metodo

di aggregazione ed il tipo di analisi, che abbiamo già detto essere – in alcuni casi – insufficienti per il nostro obiettivo, ci sono due elementi chiave che mancano agli strumenti della letteratura: un primo elemento, che possiamo trarre dall'analisi di questi strumenti, è che nessuno integra nella valutazione della camminabilità l'interferenza generata dagli attraversamenti alle intersezioni, che vedremo essere un importante elemento di disturbo della camminabilità. Il secondo elemento è che nessuno degli strumenti presentati riesce a cogliere gli aspetti teorici di base in questo testo, tra quelli proposti l'indice del *Future City Laboratory* è quello che più si avvicina ai requisiti ma manca della valutazione degli aspetti metodologici della teoria delle capacità.

3.2 Una misura della camminabilità

In riferimento a questi aspetti appena discussi, il nostro modello deve essere basato su una *network analysis* (a gravità) e deve essere costituito da un'elevata scala di dettaglio (*micro-scala*) che indaghi le caratteristiche dello spazio a livello di arredo urbano, della sua struttura, della quantità e della qualità dell'ambiente (questi aspetti costituiscono gli indicatori dei modelli). Il metodo di aggregazione dei dati e la calibrazione del modello sono gli aspetti indagati nel prossimo capitolo (il 4°) e costituiscono l'elemento di indagine fondamentale di questo lavoro.

Per cogliere gli aspetti teorici dell'approccio delle capacità è necessario misurare non solo la disponibilità delle risorse o le distanze (*commodities*) ma anche le percezioni che gli individui hanno dello spazio in funzione della sua qualità. In questo modo il livello di camminabilità di uno spazio è ottenuto misurando (1) la presenza delle opportunità urbane disponibili (*commodities*) (luoghi di interesse come parchi, negozi, servizi, ecc.) (2) la qualità (percepita) dello spazio e (3) la distanza per raggiungere le opportunità.

Questi tre elementi stanno quindi alla base della formula di misura della camminabilità che questo testo ha lo scopo di validare.

Immaginiamo quindi di aver ottenuto un punteggio di qualità percepita dello spazio pedonale (che è ciò che faremo nel prossimo capitolo), ciò che ci serve è una formula che ci permetta di aggregare distanze e opportunità urbane.

Possiamo farlo in due modi e – a scopo esemplificativo – per due finalità⁴⁵:

- Possiamo applicare un metodo compensativo⁴⁶, di somma delle caratteristiche dello spazio e della sua qualità con le distanze tra punti (cioè il valore di accessibilità). Si può, in questo modo, scegliere il percorso migliore in funzione dell'utilità che un individuo deriva dai possibili percorsi.
- Oppure possiamo adottare un metodo non-compensativo, costituito da regole di decisione annidate. A partire da una valutazione di qualità dei segmenti stradali, dai quali derivare la qualità dei percorsi, si possono definire mappe della camminabilità di punti dello spazio verso le opportunità urbane.

Ad esempio, con il primo metodo, si potrebbe determinare il miglior percorso, misurando il punteggio di camminabilità secondo una funzione di decadimento che stima il costo dello spostamento, correggendo il valore non solo in funzione delle distanze ma anche della qualità dello spazio e delle disponibilità urbane. Significa che la camminabilità c di un segmento è ottenuta da:

$$c = c_0 + \sum_{k=1}^p l_k \left(1 - \left(\sum_{l=1}^r w_l a_{kl}^r \right)^{\frac{1}{r}} \right)$$

dove c_0 è un costo fisso, l_k è la lunghezza del k -esimo segmento, a_{kl} il valore dell' l -esimo attributo di k , w_l il peso dell'attributo (per cui $\sum w_l = 1$) ed r un parametro per cui $1/(1-r)$ che indica l'elasticità di sostituzione tra gli attributi.

Se assumiamo che – secondo una concezione simile a quella del prisma spazio-temporale – un individuo che si trova in un punto dello spazio deriva un'utilità U

⁴⁵ Per maggiori dettagli e per vedere degli esempi applicativi delle due possibili procedure si veda Blečić *et al.* (2013, 2014a, 2015b) dai quali questi esempi sono tratti.

⁴⁶ Per approfondimenti sui metodi compensativi e non-compensativi si veda paragrafo 4.1.3.

in funzione della fruizione di una destinazione, possiamo calcolare la sua utilità secondo la formula:

$$U = \left(\sum_{i=1}^n X_i^\rho \right)^{\frac{1}{\rho}}$$

dove n indica il numero di destinazioni disponibili, X_i il numero di volte che l'individuo raggiunge la destinazione e dove $1/(1-\rho)$ è l'elasticità di sostituzione tra le destinazioni. Il numero di volte che l'individuo raggiunge la destinazione è influenzato dalle sue capacità intrinseche (abilità e opportunità) ed è funzione del costo c_i per raggiungere la destinazione e del budget disponibile M , tale che:

$$\sum_{i=1}^n c_i X_i \leq M$$

sotto questo vincolo, l'utilità U è massimizzata quando:

$$X_i = \frac{c_i^{\frac{1}{\rho-1}} M}{\sum_{j=1}^n c_j^{\frac{\rho}{\rho-1}}}$$

dove c_j è il costo delle altre possibili destinazioni. Il percorso migliore sarà perciò quello dal quale si deriva l'utilità massima che è funzione della distanza l_k , delle caratteristiche dello spazio $w_{i|a|k|l}$ e delle destinazioni disponibili i .

In questo caso il nostro obiettivo sarà quello di individuare i valori dei pesi w e dei coefficienti a per ogni attributo, oltre che identificare quali sono gli attributi dello spazio l percepiti come significativi nella valutazione delle camminabilità.

Il livello di camminabilità potrebbe tuttavia essere stimato a partire da un set di procedure di classificazione annidate, secondo una logica non-compensativa. Si potrebbe definire un punteggio di qualità dello spazio per ogni segmento a partire

dalle sue caratteristiche attraverso una procedura di classificazione (calibrata sulle percezioni), per poi aggregare i giudizi in percorsi tenendo conto della distanza e della disponibilità delle opportunità urbane.

Più nel dettaglio, se consideriamo ogni opportunità urbana come un punto di destinazione e definiamo dei punti di origine su una griglia $N \times N$ ⁴⁷, possiamo valutare la camminabilità di ogni origine in funzione delle classi di qualità dei percorsi, che sono a loro volta valutati dalle caratteristiche di ogni segmento che li compone. In termini più schematici significa procedere secondo quattro punti:

- Assegnare un giudizio di camminabilità ad ogni segmento e attraversamento in funzione delle sue caratteristiche qualitative e quantitative.
- Aggregare i giudizi dei segmenti di ogni percorso ottenuti nel primo *step*.
- Combinare il giudizio di ogni segmento con quelli degli attraversamenti e con la lunghezza del percorso.
- Combinare l'accessibilità delle destinazioni raggiungibili dal punto di origine, per assegnare un punteggio/classe per ogni genere di opportunità urbana.

In questo caso lo scopo del prossimo capitolo è di definire la miglior procedura di classificazione dei segmenti e attraversamenti a partire dalle percezioni degli individui. Inoltre, una breve proposta di analisi del terzo *step* è presentata nel paragrafo 4.4.

Discuteremo più avanti le possibili applicazioni dei modelli (paragrafo 4.5), ma a questo punto è utile illustrare alcuni possibili risultati che possiamo ottenere dall'applicazione di questi metodi. Possiamo ad esempio misurare la camminabilità di una città per confrontarne il livello per i diversi quartieri. Possiamo farlo in termini generali, oppure possiamo incentrare l'analisi sui diversi tipi di opportunità urbane o sulle diverse popolazioni. Gli *output* visivi saranno perciò mappe di densità della camminabilità ottenute interpolando i valori di accessibilità di ogni punto dello spazio (punti di origine), così come mostrato negli esempi applicativi in Figura 3.1.

⁴⁷ Per approfondimenti sulle griglie dei punti di origine si veda il paragrafo 2.5.3 sulla *network analysis*.

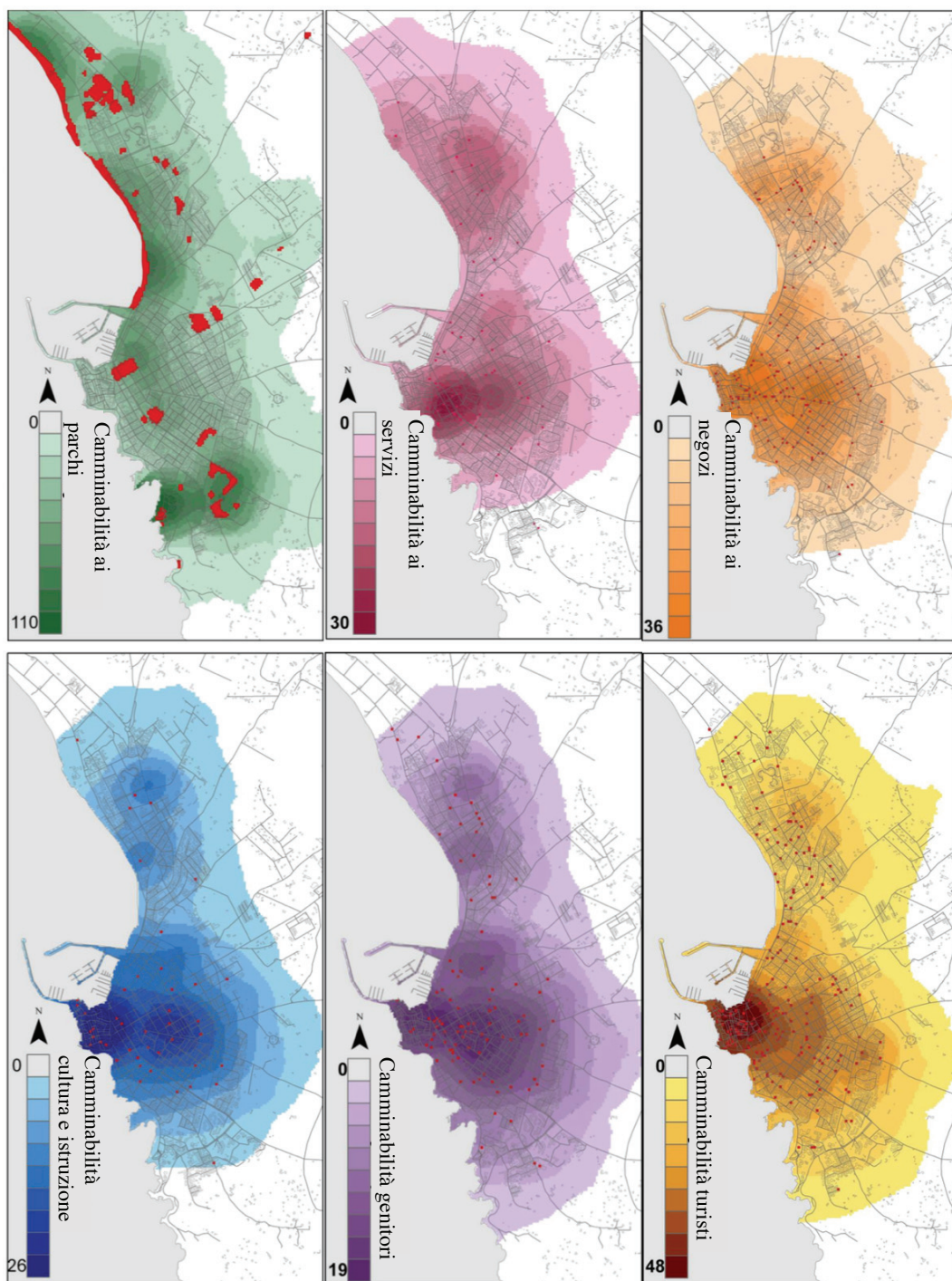


Fig. 3.1 Mappe di accessibilità per le diverse opportunità urbane e per popolazioni urbane. Fonte Blečić *et al.* (2015b).

3.3 Percezione e caratteristiche dello spazio

A questo punto il nostro scopo è duplice: da una parte identificare una procedura che ci permetta di individuare quali sono gli aspetti dello spazio che influenzano la camminabilità degli individui (e che peso hanno nella valutazione). Dall'altra calibrare la procedura di misura della camminabilità che abbiamo appena visto, testando più metodi che a partire da esempi di assegnazione ci permettano di stimare la qualità percepita dello spazio.

Lo faremo attraverso test empirici sia su segmenti stradali (tratti omogenei di strade) che su attraversamenti, che possono essere intesi come un ostacolo alla camminabilità⁴⁸; una volta identificati gli attributi significativi che definiscono i parametri del modello, andremo a testare alcune differenti procedure di classificazione che, confrontate con una metrica comune, ci permetteranno di definire il miglior modello per predire la qualità dello spazio a partire dalle sue caratteristiche. Ma non è ancora il momento di farlo, ce ne occuperemo nel prossimo capitolo (il 4°), prima è utile uno sguardo allo sfondo metodologico applicato agli studi che tentano di mettere in relazione aspetti di percezione e uso dello spazio con le sue caratteristiche, ci serve per avere una chiara visione di come strutturare i test e i modelli di previsione.

Della letteratura ci interessa analizzare più aspetti, questi sono racchiusi nei tre sottoparagrafi che seguono. I punti su cui è utile porre l'accento sono:

- Qual è l'elemento chiave della misurazione, ciò che nella pratica è la domanda da porre agli intervistati o l'aspetto del comportamento che si osserva. La scelta dell'elemento da studiare diventa fondamentale per cogliere esattamente l'obiettivo della ricerca, nella pratica questi dati corrispondono alle variabili dipendenti.
- Quali sono gli aspetti dell'ambiente urbano che nei test di relazione con l'uso e con la percezione dello spazio (variabili dipendenti) sono intesi come variabili indipendenti, in particolare quale è il livello di soggettività delle variabili e quale è la loro "scala".

⁴⁸ Il perché è discusso nel paragrafo 4.2.

- Quali sono i metodi d'indagine, ognuno dei quali ha vantaggi e svantaggi e migliori capacità di cogliere aspetti importanti nella valutazione; si tratta nella pratica dei metodi di campionamento e analisi dei dati.
- Quali sono i risultati dei test simili, che ci aiutano a capire se le conclusioni delle nostre prove sono in linea con la letteratura, in particolare rispetto alla significatività delle variabili e alla bontà di adattamento dei modelli descrittivi.

Per definire quali sono gli elementi importanti nella misura della camminabilità è utile partire dall'analisi dei fattori che influenzano la decisione del pedone e quindi la scelta del percorso. Questi fattori possono essere catalogati in caratteristiche socio-demografiche (età, reddito, genere, ecc.), caratteristiche di viaggio (motivo, frequenza, ecc.) e attributi dell'ambiente pedonale (Mateo-Babiano, 2016). Nel nostro campo di azione – quello di questa tesi – i primi due gruppi di variabili sono fissi e non possono essere modificati, mentre il terzo gruppo, le caratteristiche dello spazio, lo sono. Il focus di questo lavoro è quindi lo studio delle variabili ambientali che influenzano i comportamenti e le percezioni che gli individui hanno *nella e della* città.

L'approccio alla misura nasce a partire dalla valutazione dei bisogni del pedone, come sostiene Mateo-Babiano (2016): “Il concetto dei bisogni del pedone si è evoluto come una risposta per creare un miglior ambiente camminabile per i suoi utenti. Lo sviluppo del concetto è guidato da due lavori, la teoria dei bisogni umani (*human needs theory*) proposta da Maslow (1954) e il livello pedonale dei servizi (*Pedestrian Level of Services*) che è stato tradizionalmente usato per la progettazione dei marciapiedi (NRC, 2010)”⁴⁹. Infatti, per misurare le componenti dello spazio che influenzano la camminabilità, nella letteratura di riferimento gli autori fanno spesso riferimento al concetto di *Level Of Service* (LOS) indicato nel *Highway Capacity Manual* (HCM) (NRC, 2010). Inizialmente sviluppato per lo studio del traffico veicolare, il concetto ha trovato implicazioni nella valutazione delle caratteristiche dello spazio pedonale, per cui “si raccomanda di valutare altri

⁴⁹ “The pedestrian needs concept evolved as a response to creating a better walking environment for its users. Two bodies of work guided the development of the concept, these are the human needs theory postulated by Maslow (1954), and the pedestrian Level of Service which has traditionally been used to design sidewalks (NRC, 2010)” (Mateo-Babiano, 2016), traduzione propria.

‘fattori ambientali’ che influenzano l’esperienza a piedi, come il confort, la comodità, la sicurezza stradale e sociale, e l’attrattività” (Khisty, 1994)⁵⁰. Ciò che manca è una specifica di cosa si intenda per “fattori ambientali” ed una definizione comune degli indicatori del LOS.

La domanda alla quale cerchiamo di dare una risposta in questo studio è quindi quali caratteristiche fisiche dello spazio urbano influenzino la camminabilità, in particolare la sua percezione (qualitativa). Non siamo i primi a farci questa domanda, gli studiosi che hanno provato a rispondere a questo quesito hanno utilizzato una serie di disegni sperimentali, differenti insiemi di variabili dipendenti e indipendenti, molti metodi di indagine e intervista, e strumenti analitici di varia natura; eccone una rassegna:

3.3.1 Variabili e domande

Variabili dipendenti:

Come variabili dipendenti possiamo trovare misure del grado di soddisfazione per l’ambiente urbano (Van Dyck *et al.* 2011; Jensen, 2007), la percezione della sua qualità (Hanák *et al.* 2015; Ewing *et al.* 2006), la percezione del LOS per il pedone nei segmenti stradali (Kang *et al.* 2013) e negli attraversamenti (Ling *et al.* 2014, Muraleetharan *et al.* 2004), la disponibilità a pagare per miglioramenti dello spazio (Ng *et al.* 2015), la facilità di attraversamento (Muraleetharan *et al.* 2005), l’accessibilità pedonale (Nakamura *et al.* 2015), la percezione di sicurezza e confort (Landis *et al.* 2001), la percezione dei genitori riguardo la sicurezza dei bambini nel percorso casa-scuola (Evers *et al.* 2014; Galliforme *et al.* 2011) e la relazione tra la scelta dei percorsi tra genitori e figli (Nasar *et al.* 2015).

Quando l’obiettivo è di acquisire misure oggettive e comportamenti osservati – piuttosto che percezioni qualitative e giudizi valutativi dichiarati – gli studiosi hanno impiegato dati sull’attività (e inattività) fisica (Cerin *et al.* 2007; Gauvin *et al.* 2005), sull’uso dei trasporti pubblici (Spears *et al.* 2013), sulle frazioni di

⁵⁰ “Is also recommended to evaluate additional ‘environmental’ factors that influence the walking experience, such as comfort, convenience, safety, security, and attractiveness” (Khisty, 1994), traduzione propria.

viaggi a piedi (Lamíquez e López-Domínguez, 2015), sulle scelta dei percorsi (Koh e Wong, 2013), sulla loro praticabilità (Guo e Loo, 2013) e sulla relazione con le caratteristiche personali (Ramezani *et al.* 2015).

Variabili indipendenti:

Tra le variabili indipendenti, assunte come predittori di camminabilità, nella letteratura si incontrano tre tipi di variabili lungo l'asse oggettivo-soggettivo: (1) le caratteristiche fisiche e funzionali dello spazio urbano; (2) le pratiche di uso dello spazio (frequenze, densità, flussi, tassi di utilizzo, ecc.); e infine (3) le percezioni individuali dello spazio (es. Noria *et al.* 2013).

- Il primo tipo di variabili comprende le caratteristiche di progettazione fisica e le funzioni dello spazio. Queste misure possono essere strettamente quantitative su scala cardinale, o qualitative generalmente espresse su scale ordinali.

Le misure quantitative sono indicative della presenza o assenza di elementi che rendono lo spazio più o meno confortevole, come le indicazioni sull'illuminazione, sulla presenza di ostacoli o barriere architettoniche, alberi, elementi di protezione, trasporto pubblico e servizi o negozi; questi ultimi sono spesso misurati anche secondo la densità. Le caratteristiche geometriche dello spazio sono ugualmente espresse su scala quantitativa, come le indicazioni sulle dimensioni della carreggiata, del marciapiede, degli attraversamenti e delle intersezioni o le distanze dal centro o da vari tipi di servizi.

Quando si misurano le caratteristiche dello spazio su scala qualitativa ci si riferisce al livello di attrattività ambientale, al livello di qualità dello spazio pedonale o dell'ambiente circostante e al livello di separazione tra pedoni-veicoli e tra pedoni-biciclette, che può però anche essere misurato dalla presenza di elementi di separazione.

- Il secondo tipo di variabili descrive i fenomeni legati all'uso dello spazio, principalmente sono considerati i valori di flussi e frequenze per veicoli, biciclette e pedoni. Dei primi due si analizza spesso la velocità rilevata o limite di legge, mentre per i pedoni un importante elemento analizzato è il tempo di attesa alle intersezioni o la durata del semaforo. Altre variabili, individuate nella letteratura di riferimento, sono indicative del livello di rumore, della densità d'uso del suolo e dell'interazione tra biciclette e pedoni.

- Infine, il terzo tipo di variabili è quello costituito dalle percezioni e reazioni allo spazio, alle sue caratteristiche e uso. Nel primo caso si tratta di variabili che misurano: percezione della capacità d'accesso nello spazio, sensazione di facilità degli spostamenti, qualità percepita dello spazio, livello di sicurezza percepita e varietà dello spazio. Nel secondo caso la percezione è relativa a: camminabilità, ciclabilità e frequenza d'uso dello spazio. La tassonomia completa e la lista delle variabili è mostrata in Tabella III. Negli esperimenti in questo testo le variabili appartengono alla prima categoria per i segmenti ed alle prime due per gli attraversamenti.

Domande:

A seconda dell'obiettivo della ricerca, quando la variabile dipendente è stimata a partire da preferenze dichiarate, in letteratura si incontrano diverse domande sintetizzabili a seconda di ciò che si chiede:

- importanza diretta dei criteri (es. *“what is the relative importance of criteria”*, *“rate the level of importance for each attribute”*, *“ask your opinion on each indicator importance”*),
- percezione di qualità (es. *“evaluate each alternative and give their assessment of sidewalk LOS”*, *“score the crosswalks”*, *“how satisfied were you as a pedestrian on the road shown?”*),
- scelta del percorso (es. *“would you choice this route?”*, *“would you like to use this street?”*),
- scelta tra due alternative (choice modeling, es. *“which one is the best?”*, *“choice the best alternative”*),
- percezione di sicurezza o difficoltà (es. *“how do you perceive crossing: difficult? safety?”*, *“how safe/comfortable you feel traveling the segment?”*) e
- sensazione di attraversamento (es. *“i would feel comfortable letting an unsupervised 8-year-old child travel along/across this street/intersection”*, *“rate your familiarity with the pedestrian”*, *“how do you feel when you use this crosswalk?”*).

Riportare la lista (semplificata) delle domande incontrate in letterature ci aiuta a capire quanto sia ampio lo spettro delle possibili ricerche e quanto la domanda influisca sull'obiettivo dell'analisi. Possiamo infatti affermare che la scelta della

domanda – nella costruzione di un disegno sperimentale – debba essere ritenuta l'elemento cardine dell'indagine. A diverse domande corrispondono infatti differenti obiettivi della misura, tra i più comuni troviamo: l'analisi della correlazione tra le caratteristiche dello spazio, la misura del senso di appartenenza alla comunità, dell'inattività fisica, della soddisfazione per il quartiere, della sicurezza percepita dai genitori per i loro figli, della scelta del percorso o la definizione della relazione d'importanza tra gli attributi dello spazio.

Nel nostro caso l'obiettivo è di misurare la percezione generale di camminabilità come base di dati per inferire l'importanza dei criteri e per la costruzione del modello di predizione. Le domande del caso studio (descritte nei paragrafi successivi) puntano a carpire la sensazione di sicurezza, comodità, piacevolezza, attrattività e attraversabilità di segmenti e attraversamenti⁵¹.

Risposte:

Le risposte – nelle indagini di preferenze dichiarate – sono prevalentemente indicate su scala *Likert*⁵², da 3 a 7 punti, solo in alcuni casi la risposta è di scelta binaria (sì, no) o di preferenze (accordo, nessuna, disaccordo). Nel nostro caso, le risposte sono su scala *Likert* a 5 punti ma con precise indicazioni per rendere la valutazione il più omogenea possibile (si veda paragrafo 4.1.1).

3.3.2 Metodi di campionamento

Per quanto riguarda la raccolta dei dati e i metodi di intervista, gli studiosi hanno intrapreso differenti procedimenti. Le procedure di intervista diretta in strada possono essere classificate in (Ling *et al.* 2014):

⁵¹ In particolare per i segmenti sono valutati sicurezza, comodità, piacevolezza, attrattività e attraversabilità ritenuti “obiettivi mediani” nella procedura di definizione delle variabili indipendenti tramite il metodo del *Value Focused Thinking* (Keeney 1992, Keeney 1996). Gli attraversamenti sono invece valutati secondo sicurezza e comodità che, secondo la letteratura, sono gli elementi d'importanza.

⁵² Per approfondimenti si veda: Likert R. (1932) *Technique for the measure of attitudes*, Arch. Psycho., Vol. 22 N. 140.

Tab. III Tassonomia delle variabili indipendenti degli studi menzionati.

Tipo variabile	Variabili
[1] Caratteristiche fisiche, funzionali e di progettazione dello spazio	SERVIZI E NEGOZI - Retail and shops density or presence, Land use type CARATTERISTICHE CARREGGIATA - Number of lanes, Center turning lane, Median width, Enclosure, Directional signs CARATTERISTICHE MARCIAPIEDE - Level of space at corner, Path obstructed, Path width, Space at corner, Continuous sidewalk, Type of pedestrian facility (sidewalk), Street amenities ATTRAVERSAMENTI - Crossing Facilities, Crossings marked, Elevated crossing for pedestrians, Refuge islands INTERSEZIONI - Corners with bump-outs, Size of intersecting roads, Traffic control device, Visible cross markings, number of road crossings, Turning vehicles delay, Number of crossings, attraversamenti rialzati, semaforo, isola di attraversamento DIFFICOLTÀ MOTORIE - Curb ramps, Steps/slopes, Traversable by wheelchair, Consider transport disadvantaged persons Needs (e.g. facilities for the blind) ILLUMINAZIONE - Daylight indicator OSTACOLI - Driveways cross path, Fences, Grade crossings PARCHI E PIAZZE - Lawn, Open space, Ends in cul-de-sac ALBERI - Plants or trees indicator, Shrubs, Street trees TRASPORTO PUBBLICO - Public transport stops and stands, subway stations PROTEZIONI - Shaded walkways, Weather protection VALORI GENERALI DELLA STRADA - Street connectivity z-score, Walkability z-score, street type
[1a] (Quantitative)	PERCORSI, DISTANZE- Distance to [different services], Distance to the city center, Route length, Length of walking time
(Qualitative)	SEPARAZIONE PEDONI-VEICOLI - separation of sidewalk from motor-vehicle traffic indicator, Lateral separation elements between pedestrians and motor vehicle traffic, Plant buffers between sidewalk and street
[1b]	SEPARAZIONE PEDONI-BICI - Separate bicycle path TIPO DI QUARTIERE - House type, House, The architectural appearance of the neighbourhood, Type of land use/buildings, CONDIZIONI MARCIAPIEDE - Debris, Decorated sidewalk surface indicator, Free of tripping hazards, Sidewalk condition, Street condition, Visual damage of sidewalk surface ATTRATTIVITÀ AMBIENTALI/CULTURALI - Cultural elements, Good scenery, Well maintained surroundings (environment), Hiding places (influenzano la percezione della sicurezza).
[2] Pratiche d'uso dello spazio	FREQUENZE AUTO - Flow rate, level of traffic, Number of vehicles encountered, Traffic volume, Motor vehicle mix (i.e., percentage of trucks), parcheggi che ostacolano il passaggio e la visuale FREQUENZE BICI - Bicycle flow rate on sidewalk, Bicyclist volume FREQUENZE PEDONI - Crowdedness, Pedestrian flow rate on sidewalk INTERAZIONE PEDONI-BICI - Bicyclist against walking direction indicator VELOCITÀ TRAFFICO VEICOLI - Average speed of motor vehicles, VELOCITÀ TRAFFICO BICI - Bicyclist average speed, Maximum bicyclist speed BICI - Bicycle events DENSITÀ - Jobs/residents, Population density, Resid+jobs+ students at schools/Ha, Retail food units/residents, Retail jobs/residents, Retail units/residents RUMORE - Noise, Noise or construction INTERSEZIONI - Police stand at intersection, Street-crossing waiting time, Waiting time, parcheggio alle intersezioni
[3] Percezioni o reazioni allo spazio individuali	ACCESSO - Accessibility, Access to services, Connectivity, Linkage, System Continuity QUALITÀ VISIVA DELLO SPAZIO - Aesthetics, Attractiveness, Pleasurability, Imageability (quality of space that makes it distinct), Tidiness (condition and cleanliness of a place), Environmental stimuli, Visual interest, Adequate Lighting, Disorder SICUREZZA - Crime safety, Traffic, Safety (walkability), Security (crime), visibilità alle intersezioni FACILITÀ - Comfort, Convenience (comodità), Human scale, Legibility (ease with which the spatial structure of a place can be understood and navigate)
[3a] Caratteristiche dello spazio	VARIABILITÀ - Complexity (frequency of space, land use), Diversity (building styles and socializing opportunities), Coherence (percezione dello spazio non conosciuto), Familiarity with environment
[3b] Uso dello spazio	CICLABILITÀ - Bicycle system addresses cyclists' needs, Bicycle system limits, Effort to cycle around CAMMINABILITÀ - Effort to walk around, Pedestrian perceived crossing difficulty and safety, Pedestrian system addresses pedestrian needs, Threat of traffic, Pedestrian limits FREQUENZE - Inclusive of people, Number of people-oriented destinations

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
 Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

- *observational method* (OM), consiste nella valutazione del LOS basata sull'osservazione *in-situ* del comportamento dei pedoni (densità pedonale, frequenza dei pedoni, velocità camminata, ecc.);
- *intercept survey* (IS), consiste nell'intervista ai pedoni dopo aver attraversato un incrocio o un segmento stradale e nella richiesta di valutazione dell'attraversamento o del segmento stradale;
- *contingent field survey* (CFS), consiste nel coinvolgere soggetti a piedi, istruiti per valutare dei segmenti o attraversamenti prestabiliti subito dopo aver percorso il tratto stradale o attraversato l'incrocio;
- *controlled field valuation* (CFV), consiste nella valutazione dei segmenti o degli attraversamenti in maniera simile alla procedura precedente ma senza richiedere ai valutatori di percorrere il segmento o attraversare l'incrocio. È generalmente adottata per gli attraversamenti, ma può essere adattata anche ai segmenti;
- *laboratory/simulation studies* (LSS), consiste nell'osservazione e valutazione di rappresentazioni dell'ambiente urbano. Le simulazioni possono comprendere varie tecniche di descrizione, rappresentazione e visualizzazione dello spazio camminabile, da rendering 3D (es. Nasa *et al.* 2015) a fotografie e fotomontaggi (es. Ng *et al.* 2015), sino a video clip (es. Jensen, 2007).

Altri approcci, usati per determinare la percezione o i comportamenti dei pedoni, sono basati su metodi standard di intervista (intervista in strada, intervista telefonica o per posta, questionario *web-based*, ecc.). I metodi di preferenze rilevate *Revealed Preference methods* (RP) sono usati per indagare i modelli di scelta del percorso (*route choice*) (es. Guo e Loo 2013), mentre i metodi di preferenze dichiarate *Stated Preferences methods* (SP) per conoscere valutazioni delle caratteristiche dello spazio (es. Cerin *et al.* 2007; Hanák *et al.* 2015; Koh e Wong, 2013; Guo e Loo, 2013).

Per il nostro scopo – carpire le percezioni degli individui come *proxy* dei loro bisogni e desideri (in riferimento alla teoria Seniana) – il CFS è lo strumento adatto ad ottenere un giudizio di qualità dello spazio. In accordo con Guo e Loo (2013) “la valutazione contingente basata sulle preferenze dichiarate può sovrastimare l'importanza degli attributi più rilevanti, come la distanza e la sicurezza, poiché i pedoni non sono spesso capaci di interpretare dotazioni

intangibili, come il paesaggio urbano o l'architettura dei luoghi"⁵³. Raccogliere i giudizi di camminabilità *in-situ* permette inoltre di cogliere il più possibile le reali percezioni dello spazio, che potrebbero perdersi con i metodi di intervista standard.

3.3.3 Test simili, metodi di analisi e risultati

Tra quelli citati, ci sono diversi studi che adottano una procedura di analisi che può essere utile confrontare per gli scopi del nostro esperimento.

Ricordiamo che il nostro obiettivo è quello di identificare quali attributi dello spazio contribuiscono alla sua valutazione e – con questi – costruire un modello di stima della qualità. Ci interessa quindi capire quali attributi sono significativi, quale procedura è stata adottata e quali sono i risultati delle analisi. Una sintesi è proposta in Tabella IV. Per una lettura più scorrevole, se il lettore possiede già informazioni sulla letteratura di riferimento, è possibile bypassare la prima parte di questo sottoparagrafo, saldando al suo ultimo blocco (pag. 74) che, ragionando sugli elementi delle procedure, trae le considerazioni sui metodi della letteratura e le differenze con il metodo proposto.

Nasar *et al.* (2015) applicando la procedura LSS hanno condotto un'analisi sulla differenza tra le scelte del percorso di genitori e figli e su come la qualità dello spazio (da loro definita *physical disorder*) possa influenzare le loro scelte. Attraverso la rappresentazione simulata dello spazio gli autori hanno potuto costruire un disegno fattoriale frazionato generando così nove combinazioni per "livello di qualità", "larghezza della strada" e "tipo di nascondigli". L'esperimento consiste nella scelta (per genitori e figli) tra coppie di strade con differenti livelli di qualità. L'analisi statistica dei risultati è condotta con un modello di regressione logistica ad effetti misti (*mixed effects logistic regression model*) per testare la probabilità di scelta di una strada (di un genitore o di un figlio) all'aumentare del

⁵³ "contingent rating based on stated preference may overestimate the importance of more tangible attributes, such as distance and safety, because pedestrians were often unable to articulate intangible amenities, such as streetscapes and façade designs" (Guo e Loo 2013), traduzione propria.

livello di qualità. A parte le principali considerazione tra le relazioni di preferenza di genitori e figli, gli autori hanno ottenuto alcune considerazioni sul come lo spazio influenza le scelte; infatti, l'analisi ha confermato l'importanza della qualità dello spazio e suggerisce che il livello di traffico, la presenza di verde e la presenza di spazi nascosti influenzano la valutazione. Analizzando le ragioni delle risposte, i ricercatori hanno inoltre individuato che immondizia abbandonata e sporcizia influenzano fortemente (e negativamente) la scelta, mentre elementi naturali (es. alberi, prati, ombra, ecc.) e sicurezza del pedone rispetto alle auto (es. poche auto, meno traffico, segnali di stop, ecc.) influenzano positivamente le percezioni.

Muraleetharan *et al.* (2004) adottando una *conjoint analysis* per determinare il *trade-off* tra attributi. Nel loro studio hanno investigato come i pedoni valutano il LOS di segmenti stradali ottenendo che la "frequenza dei pedoni" è l'attributo più importante (30%), la "frequenza delle biciclette" e la "larghezza e separazione" sono in seconda battuta (28%), mentre la presenza di "ostacoli" ha un'importanza relativa del 15%. Altri impatti di utilità positiva sono registrati per "oltre 3 metri di larghezza e eccellente separazione", "nessun ostacolo" e "bassa frequenza di biciclette"; mentre bassi livelli sono registrati per "meno di 1,5 metri e nessuna separazione", "oltre 5 ostacoli per 100 metri" e "alta frequenza di biciclette". Per gli attraversamenti hanno ottenuto che la "direzione dei veicoli" è l'attributo più importante (32%), lo "spazio agli angoli delle intersezioni" è il secondo (26%), il "tempo di attesa per l'attraversamento" e le "infrastrutture per l'attraversamento" sono al terzo (21%). In particolare, valori di utilità elevati dei livelli delle variabili sono registrati per "nessuna svolta dei veicoli" e "spazio di sosta e circolazione abbondante agli incroci"; parallelamente, la "svolta dei veicoli sia da destra che da sinistra verso l'attraversamento" genera bassi livelli di utilità.

Kadali e Vedagiri (2015), attraverso un'indagine con questionari e video simulazioni, hanno valutato la percezione della difficoltà (e di sicurezza) nell'attraversamento in relazione al tipo di zona e alle caratteristiche geometriche dell'incrocio. Adottando un modello di regressione ordinale probabilistico (*ordered probability model*) hanno testato la relazione tra il LOS e le variabili indipendenti. Il fattore più importante risulta essere il numero di veicoli ed un

aumento della frequenza di veicoli sembra influenzare significativamente i comportamenti pedonali.

Nakamura (2016) ha testato quanto l'accessibilità pedonale dipende dalle caratteristiche dello spazio su specifiche destinazioni. Il modello adottato mette in relazione la frequenza (osservata) dei pedoni – come variabile dipendente – con le caratteristiche dello spazio. L'autore applica una regressione *stepwise*, le caratteristiche dello spazio su diverse scale sono gradualmente aggiunte al modello per identificare l'influenza che hanno rispetto alla frequenza dei pedoni. Del campione composto da 116 segmenti stradali sono stati registrati “larghezza del marciapiede” e “grado di copertura di negozi”, che in relazione alla variabile dipendente ottengono un R-quadro=0,56. Altri due modelli “percorso nel quartiere” e “percorso fuori dal quartiere” analizzati con le variabili “numero di strade”, “distanza” e “traffico trasversale” ottengono rispettivamente un R-quadro di 0,64 e 0,63. I risultati principali di questo studio mostrano che i pedoni sono meno disposti a camminare per lunghe distanze, ma piuttosto preferiscono brevi spostamenti in strade pedonali. L'attrazione dello spazio, così come la qualità della pavimentazione contribuisce molto di più nei percorsi brevi che in quelli più lunghi.

Guo e Loo (2013) propongono un test per quantificare l'utilità percepita nell'ambiente pedonale e per testare un modello di scelta del percorso (*route choice modelling*). L'analisi conta un campione di 321 pedoni tra New York (NYC) e Hong Kong (HK). Dall'applicazione di un modello probabilistico multinomiale (*multinomial probit model*), risulta che le uniche correlazioni moderate sono tra il numero di attraversamento e la lunghezza dei percorsi per entrambi i campioni ($r=0,65$), mentre il volume di traffico e la presenza dei negozi ha una correlazione di $r=0,67$ nel campione per HK. Nel modello per la città di NYC, le due variabili significative sono la presenza di negozi e spazi aperti. Per di più, con segno atteso, la lunghezza del percorso e la larghezza del marciapiede sono significativi ad un livello del 10%. Mentre per il modello di HK risulta principalmente significativa la lunghezza del percorso. Gli autori trovano che il modello HK tende ad essere meno preciso nella stima del giudizio finale, spiegando questo a causa della maggiore complessità dell'ambiente di HK.

Mateo-Babiano (2016) esamina il processo decisionale dei pedoni applicando l'*Analytic Hierarchy Process* (AHP)⁵⁴ per definire empiricamente la gerarchia delle esigenze dei pedoni (criteri) e per esaminare le priorità relative degli attributi ambientali (alternative) che soddisfano i bisogni dei pedoni. Agli intervistati è stato chiesto di esprimere la loro preferenza (su scala Saaty) tra confronti a coppie di 12 attributi dello spazio. I risultati indicano che la consistenza dei giudizi è buona $CI/RI < 0,10$ ⁵⁵ e che la gerarchia delle necessità è: protezione, facilità, equità, mobilità, identità e godimento.

Kang *et al.* (2013) hanno investigato la percezione del LOS dei pedoni rispetto alle caratteristiche dello spazio. Per l'analisi è stato applicato un modello ordinale probabilistico (*ordered probability approach*), il primo risultato è che l'impatto della larghezza del marciapiede è il fattore principale, così come la frequenza dei pedoni. Il secondo risultato è che tutti gli indicatori legati alle biciclette hanno un forte impatto negativo. Terzo, la presenza di negozi, la separazione tra pedoni e veicoli e la presenza di parcheggi hanno una forte influenza con la percezione del LOS.

Landis *et al.* (2001) hanno testato un metodo per quantificare oggettivamente la percezione dei pedoni del livello di sicurezza e confort. Applicando una regressione *stepwise* su 1250 osservazioni (di 75 individui) hanno analizzato la relazione tra le due percezioni espresse su scala *Likert* a 6 punti e le caratteristiche dello spazio: larghezza del marciapiede, separazione tra pedoni e veicoli, parcheggi su strada, larghezza della corsia di marcia e della pista ciclabile, frequenza e velocità auto, tipologia veicoli e frequenza e volume di accessi carrai. Un modello di stima finale con 11 variabili produce un R-quadro = 0,85.

Koh e Wong (2013) hanno condotto un'indagine per esaminare l'influenza dei fattori delle infrastrutture sulle scelte di spostamento di pedoni e ciclisti. I ricercatori hanno condotto interviste sulle preferenze dichiarate con *audit* dello

⁵⁴ Per approfondimenti sul modello AHP si veda Saaty, T.L. (1980) *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw Hill, New York. O in versione ridotta Saaty, T.L. (2008) Decision making with the analytic hierarchy process, *Int. J. Services Sciences*, 1(1): 83-98.

⁵⁵ Consistency index (CI) e random index (RI). Per approfondimenti si veda Saaty (2008). In breve, indicano quanto, nel processo di classificazione delle alternative, le risposte degli intervistati sono attendibili secondo la proprietà di transitività: $A > B \wedge B > C \Rightarrow A > C$.

spazio pedonale e ciclabile, ed hanno comparato le scelte del percorso con 11 fattori dello spazio (pedonale). Lo studio ha rilevato una preferenza per le strade con un elevato grado di confort, con negozi, con buoni scenari e preferibilmente con un elevato affollamento.

Lamíquiz e López-Domínguez (2015) applicano un modello di regressione multivariato (*multivariate regression model*) per studiare l'associazione tra le caratteristiche dello spazio urbano e la proporzione di pedoni in diverse parti della città. Le variabili indipendenti sono state organizzate in tre gruppi: (1) *network* di strade (lunghezza percorso, lunghezza tratti, ecc.) e la sua configurazione di accessibilità (connettività, integrazione, ecc.); (2) tipo di uso del suolo (densità, mixité, ecc.); e (3) variabili dello spazio non costruito (caratteristiche socio-demografiche, come età o possesso di un'auto). Il miglior modello di regressione ottiene un R-quadro (*adjusted*) del 70,63% per 14 variabili delle tre categorie.

Evers *et al.* (2014) hanno effettuato uno studio su strade e intersezioni principali per sette scuole elementari in un quartiere suburbano. Ai genitori degli studenti di quelle scuole è stato chiesto di esprimere il loro accordo con la locuzione “*I would feel comfortable letting an unsupervised 8-year-old child travel along/across this street/intersection*” su una scala *Likert* a 5 punti centrata su neutrale. Un modello di regressione logistica (*logit*) ottiene R-quadro=0,63 e indica una associazione significativa tra le sei variabili: piacevolezza nel camminare, attraversamenti per disabili, dimensione delle intersezioni, ostacoli, alberi e assenza di pericoli.

Ling *et al.* (2014) hanno valutato la percezione del LOS applicando un'analisi di correlazione e una regressione *stepwise*. Lo studio analizza come la percezione di sicurezza e confort nell'attraversamento (rilevata tramite CFS) sia influenzata da: svolta dei veicoli da destra, uscita di biciclette da sinistra, affollamento pedonale, affollamento ciclisti, tempo di attesa di attraversamento, presenza dell'isola di attraversamento e attraversamento in due *step*. La bontà di adattamento R-quadro del modello è 0,65.

Muraleetharan *et al.* (2005) hanno provato a identificare i fattori che influenzano il LOS pedonale per le intersezioni. Il modello adottato per l'analisi è una regressione *stepwise*, che ha dimostrato che il fattore “svolta dei veicoli” ha l'influenza maggiore nella percezione del LOS. I risultati mostrano che quando il numero di veicoli in svolta aumenta c'è una corrispondente decrescita della

sicurezza percepita. Inoltre, i fattori “tempo di attesa per l’attraversamento” e “interazione tra pedoni e ciclisti” sono stati individuati come fattori determinanti del LOS.

Jensen (2007) ha cercato di determinare le variabili chiave che influenzano la soddisfazione dei pedoni (preferenze dichiarate). Il modello di regressione finale del LOS produce un valore di R-quadro di 0,55 utilizzando come variabili indipendenti: “frequenza veicoli”, “velocità media dei veicoli a motore”, “tipo di struttura pedonale (marciapiede - senza marciapiede)”, “tipo di struttura per biciclette (pista ciclabile a una via - pista ciclabile due vie - pista su strada)” e “il tipo di uso del suolo / edifici (shopping - residenziale - uso misto)”.

Muraleetharan *et al.* (2005) hanno testato anche un metodo per valutare il LOS per attraversamenti alle intersezioni tramite una regressione *stepwise*, comparando il giudizio di percezione delle intersezione rispetto a caratteristiche geometriche, d’uso e di traffico. Nello studio sono analizzati da 252 partecipanti 17 attraversamenti in 4 differenti intersezioni. Nel modello finale il giudizio è influenzato dagli attributi: “spazio di sosta e passaggio alle intersezioni”, “elementi che facilitano l’attraversamento”, “direzione dei veicoli in svolta”, “tempo di attesa per l’attraversamento” e “numero di interazioni tra pedoni e ciclisti”.

Ling *et al.* (2014) applicano una CFS e una LSS video per stimare il LOS pedonale basato sull’ambiente urbano cinese. I fattori che influenzano la qualità dello spazio sono stati investigati con una regressione *stepwise* per cui R-quadro=0,65. La ricerca, condotta in 12 attraversamenti (in 6 incroci) ha un totale di 383 osservazioni sperimentali. I fattori che dal modello di regressione hanno un’influenza negativa sono: “frequenza veicoli in svolta”, “frequenza traffico a due ruote” e “tempo di attesa per l’attraversamento”. Mentre “affollamento pedoni”, “presenza di isole di attraversamento” e “attraversamenti in due *step*” hanno un effetto positivo nel valore di LOS.

La Tabella IV riassume gli studi menzionati, specificando l’obiettivo principale, il metodo di investigazione (in riferimento alla classificazione precedente dei metodi di indagine, paragrafo 3.3.2), il tipo e il numero di variabili (secondo la tassonomia in Tabella III), la domanda proposta e il tipo di risposte ammesse, il

metodo di analisi dei dati e la numerosità campionaria. Rispetto ai casi della letteratura citata, l'analisi condotta in questo testo – descritta a partire dal prossimo paragrafo e capitolo – esamina in media un numero di variabili maggiore e lo fa con un livello di dettaglio superiore ai casi proposti per tre motivi: (1) si indaga l'ambiente urbano su una *micro*-scala che analizza la presenza dell'arredo urbano, (2) si esaminano un numero di caratteristiche (variabili) dello spazio maggiori e (3) queste sono descritte con maggior dettaglio, cioè con un elevato numero di modalità⁵⁶. Per l'interpretazione delle variabili sono usate descrizioni sia qualitative che quantitative per ridurre equivoci ed errori di interpretazione della scala di misura. Il numero del campione esaminato è generalmente in linea con la letteratura, così come le procedure di regressione, i suoi risultati ($R\text{-quadro} \approx 55\div 70\%$) e la scala di misura della variabile dipendente. L'elemento principale di vantaggio rispetto ai casi della letteratura è che, per la costruzione del modello di previsione della qualità dello spazio, sono adottati non solo strumenti compensativi derivati dai test di regressione, ma nella analisi sono inclusi modelli non-compensativi basati su altre procedure di previsione.

Tab. IV Sintesi degli studi sulla camminabilità e la relazione con le caratteristiche dello spazio.

Obiettivo	Autori	Metodo raccolta dati	[tipi di variabili] (N°)	Domanda	Risposta	Metodo/ Modello	Numerosità campione
Relazione tra qualità ambiente e scelta percorsi	Nasar <i>et al.</i> 2015	LSS (3D)	[1a, 1b] (14)	<i>Choice experiment (Pairwise comparison)</i>	7 punti scala Saaty	<i>Logistic regression</i>	27 (scenari)
Relazione tra caratteristiche dello spazio e spostamenti pedonali	Lamíquiz <i>et al.</i> 2015	-	[1a, 2] (20)	<i>Percentage of walking trips (observation)</i>	-	<i>Bivariate correlation</i> <i>Multiple regression</i>	-
Relazione tra le caratteristiche dello spazio e la percezione di sicurezza dei genitori	Evers <i>et al.</i> 2014	CFV	[1a] (11)	<i>I would feel comfortable letting an unsupervised 8-year-old child travel along/ across this street/ intersection</i>	5 punti scala Likert	<i>Logistic regression</i>	1641 (segmenti stradali classificati)
Definizione dei pesi attributi	Muraleeth aran <i>et al.</i> 2004	LSS (Draws)	[1a, 2] (7)	<i>Would you like to use this street?</i>	10 punti scala Saaty	<i>Conjoint Analysis</i>	531 (questionari)
Valutazione del LOS alle intersezioni	Kadali <i>et al.</i> 2015	LSS (Video)	[1a, 3a] (9)	<i>How do you perceive crossing:</i> - <i>difficult?</i> - <i>safety?</i> - <i>LOS?</i>	6 punti scala Likert	<i>Ordered probability model</i>	1040 (questionari) in 8 intersezioni

⁵⁶ La modalità statistica indica i possibili valori di una variabile discreta (chiamata anche carattere).

Tab. IV (continua)

Obiettivo	Autori	Metodo raccolta dati	[tipi di variabili] (N°)	Domanda	Risposta	Metodo/ Modello	Numerosità campione
Relazione tra caratteristiche dello spazio e accessibilità	Nakamura 2016	OM	[1a] (5)	-	-	<i>Stepwise regression</i>	116 (segmenti stradali)
Scelta dei percorsi pedonali	Guo <i>et al.</i> 2013	RP	[1a, 2] (6)	-	-	<i>Probit regression</i>	321 (questionari)
Relazione tra caratteristiche dello spazio e percezione	Hanák <i>et al.</i> 2015	SP	[1a, 1b, 2, 3a] (22)	<i>Ask your opinion on each indicator importance</i>	5 punti scala Likert	<i>Regression analysis</i> (tra criteri)	595 (questionari)
Relazione tra caratteristiche dello spazio e LOS	Muraleetharan <i>et al.</i> (2005)	OM	[1a, 1b, 2] (9)	<i>How do you feel when you use this crosswalk?</i>	10 punti scala Saaty	<i>Stepwise regression</i>	252 (questionari) in 4 intersezioni
Valutazione della soddisfazione dichiarata	Jensen 2007	LSS (Video)	[1a, 2] (5)	<i>How satisfied were you as a pedestrian on the road shown?</i>	6 punti scala Likert	<i>Logit regression</i>	7724 (segmenti stradali) di 56 vie
Gerarchia dei bisogni dei pedoni	Mateo-Babiano 2016	SP	[1a, 1b, 2] (20)	<i>Choice between two criteria</i>	7 punti scala Saaty	AHP	70 (questionari)
Come i pedoni percepiscono il LOS	Kang <i>et al.</i> 2013	LSS (Video)	[1a, 1b, 2] (11)	<i>Evaluate each alternative and give their assessment of sidewalk LOS</i>	6 punti scala Likert	<i>Ordered probability model</i>	114 (questionari)*15 (scenari)
Percezione dei pedoni della sicurezza e del confort	Landis <i>et al.</i> 2001	CFS	[1b, 2] (4)	<i>how safe/comfortable you feel traveling the segment?</i>	6 punti scala Likert	<i>Stepwise regression</i>	1250 (osservazioni) di 75 individui in 21 segmenti stradali
Relazione tra caratteristiche dello spazio e scelta dei percorsi	Koh <i>et al.</i> 2013	SP	[1a, 1b, 3a] (9)	<i>Rate the level of importance for each attribute</i>	4 punti scala Likert	<i>t-tests</i> (tra valori medi d'importanza degli attributi)	1146 (questionari) in 5 stazioni treni
Come i pedoni percepiscono il LOS	Ling <i>et al.</i> 2014	CFS	[1a, 1b, 2] (10)	<i>Score the crosswalks</i>	5 punti scala Likert	<i>Stepwise regression</i>	383 (questionari) in 12 intersezioni

Nota: La "Saaty scale" indica una modalità di comparazione paritaria (confronto a coppie) definita nell'AHP, mentre la "Likert scale" è una scala ordinale di gradimento.

Capitolo 4

Test empirici e modelli predittivi

Da questo punto il testo è dedicato alla descrizione degli esperimenti condotti ed ai modelli predittivi testati. Prima di introdurli è utile una precisazione sulla scelta di alcuni degli elementi teorici discussi in precedenza e degli elementi metodologici analizzati nel capitolo appena concluso, come il metodo di analisi e la scelta delle variabili. Da un punto di vista teorico gli esperimenti costituiscono uno strumento della psicometria (e modellizzazione delle preferenze) utile ad intercettare gli elementi ritenuti importanti nella valutazione della qualità dello spazio in funzione della *capacità* e dell'interesse d'uso degli individui. Applicare il concetto di capacità urbana nella misura di camminabilità significa non contabilizzare solo disposizioni urbane (*commodities*), distanze, disponibilità e distribuzione delle risorse, ma significa valutare i funzionamenti (*functionings*) che possiamo intercettare dalle percezioni. In questo modo, è possibile derivare dai giudizi di percezione che gli individui hanno dello spazio ciò che è da loro ritenuto più influente nella valutazione della qualità dello spazio, come pure quanto l'ambiente urbano è camminabile ed attraversabile.

Lo scopo di questo capitolo è quindi quello (1) di testare alcune procedure di inferenza della qualità dello spazio a partire delle sue caratteristiche, attraverso gli strumenti della psicometria e (2) di produrre modelli di previsione costruiti con metodi di supporto alla decisione dell'analisi multicriteriale e di *supervised machine learning*.

Per definire quali sono gli elementi dello spazio importanti nella valutazione sono stati adottati due esperimenti, effettuati su segmenti stradali e – in più rispetto ai modelli di camminabilità della letteratura – sulle intersezioni. Si basano sull'indagine della qualità percepita della camminabilità e del rapporto tra le caratteristiche dello spazio. La domanda è quindi costruita in modo che gli intervistati forniscano un giudizio in funzione di sicurezza, comodità, piacevolezza, attrattività ed attraversabilità percepita; per questo compito il metodo che meglio permette di cogliere le percezioni è il CFS. Le variabili – come anticipato più volte – sono espresse su una *micro*-scala di dettaglio,

attraverso una descrizione puntuale degli elementi di arredo urbano e caratteristiche dell'ambiente (su scala sia quantitativa che qualitativa) e sulle pratiche d'uso dello spazio; in riferimento agli studi citati quello proposto ha il vantaggio di considerare un numero di aspetti dello spazio (variabili) ed un dettaglio (modalità) maggiore. La relazione tra le variabili è indagata servendosi di vari test di regressione, secondo l'idea di poter ricavare informazioni sulla relazione di causa-effetto tra le caratteristiche dello spazio ed il giudizio soggettivo di qualità percepita.

Il secondo obiettivo degli esperimenti è di definire una procedura di classificazione automatica. Per compensare l'assenza in letteratura, i modelli testati sono sia compensativi che non-compensativi. Dei primi sono testate la regressione lineare multipla, la regressione *stepwise* e la regressione ordinale. Della seconda categoria, i metodi applicati sono: il *dominance-based rough set approach* che definisce regole di assegnazione del tipo *if...then*; il metodo *Majority Rule Sort* (e ELECTRE TRI) che classifica le alternative secondo una regola di maggioranza ed infine le reti neurali artificiali che classificano le alternative attraverso percorsi non lineari non definiti a priori.

Nei paragrafi 4.1 e 4.2 sono descritti i test di regressione e i modelli di previsione, rispettivamente per i segmenti stradali e per gli attraversamenti. Nel paragrafo 4.3 si propone un confronto sui risultati dei modelli, che include alcuni test di verifica del livello di sovra-adattamento dei modelli e la valutazione dei risultati rispetto ad un modello di previsione casuale. Nel paragrafo 4.4 sono proposti alcuni possibili strumenti di analisi delle relazioni di scelta tra segmenti, attraversamenti e distanze. Infine, il paragrafo 4.5 è destinato alle riflessioni sui possibili usi dei modelli e dello strumento di misura della camminabilità.

4.1 Segmenti stradali

Per segmenti stradali si intende una sezione omogenea di strada composta da determinate caratteristiche, una strada è divisa in più segmenti quando le sue caratteristiche cambiano. I segmenti compongono quindi la base della misura e – insieme agli attraversamenti – sono l'elemento della misurazione di camminabilità

dello spazio. La loro analisi, in relazione alle percezioni di qualità dello spazio, è descritta nei sottoparagrafi seguenti.

4.1.1 Costruzione dell'esperimento

Sulla base degli studi appena elencati ed in accordo con i nostri requisiti di ricerca discussi in precedenza, è stata condotta una *contingent field survey* CFS per l'intera rete stradale della città di Alghero. Lo scopo della nostra indagine è quello di raccogliere due misure per ogni segmento stradale⁵⁷ della città: (1) una descrizione analitica del segmento stradale, mediante 18 attributi⁵⁸ della strada, collezionati tramite un *audit*, e (2) una valutazione sintetica soggettiva di percezione della qualità dello spazio pedonale, o per meglio dire di camminabilità. La rete stradale di Alghero è stata divisa in 408 segmenti omogenei e l'intera città in 10 settori (Figura 4.1). I dati sono stati raccolti nel Gennaio 2016 da 24 studenti suddivisi in 12 gruppi, ogni coppia ha condotto (1) un *audit* per la raccolta dei 18 attributi di un'area ed ha fornito (2) una valutazione sintetica personale di qualità dello spazio di un'altra area. Ogni gruppo ha raccolto dati per aree differenti per non avere una conoscenza analitica dettagliata che potesse influenzare fortemente la valutazione sintetica delle strade.

In totale sono state raccolte valutazioni per 483 segmenti, ridotti a 408, uno per ogni segmento stradale calcolando la "media" dei valori di più record che si riferiscono allo stesso segmento, correggendo poi i dati per avere variabili ordinali (ovvero numeri naturali).

Le caratteristiche dello spazio da analizzare sono definite a partire da studi preliminari (Blečić, *et al.* 2015a) e secondo i riferimenti della letteratura; sono state inoltre incrementate con l'uso della procedura del *Value Focused Thinking* proposta da Keeney (1992, 1996).

⁵⁷ Secondo la definizione di segmento stradale ogni strada della città è stata divisa in più parti in funzione delle sue caratteristiche per un totale di 408 segmenti.

⁵⁸ Il numero delle caratteristiche inizialmente analizzate è maggiore, questa proposta è una riduzione che esclude le variabili con errori di misurazione o variabilità insufficiente per l'analisi. Le variabili escluse sono: densità edifici, pendenza, presenza parchi e aree sportive, fermate di trasporto pubblico o condiviso, presenza di muri o ferrovie che costeggiano la via.



Fig. 4.1 10 settori della città (destra), ed esempio di segmenti stradali omogenei per il settore 7 (sinistra).

I livelli di ogni indicatore sono espressi in forma qualitativa ordinale su scala da 1 (situazione migliore) a 2, 3, 4 o 5 (situazione peggiore). La lista delle 18 caratteristiche oggettive dello spazio è riportata in Tabella V.

Mentre alcuni autori tra quelli citati adottano variabili qualitative scalari, in questo studio ogni variabile è descritta su scala ordinale che ne descrive le caratteristiche qualitative e quantitative. Ogni livello (modalità) delle caratteristiche dello spazio (variabili indipendenti) è descritto ai revisori in maniera dettagliata e per mezzo di esemplificazioni, per tentare di limitare ambiguità e interpretazioni soggettive. Ad esempio, l'attributo "larghezza del marciapiede" è composto dai cinque livelli: Ampio, Comodo, Minimo, Inadeguato e Assente. Secondo i differenti sistemi di stima degli intervistati, queste descrizioni potrebbero essere dissimili, per questo – nelle linee guida fornite per l'*audit* – ogni livello è stato descritto accuratamente. Quindi, per la larghezza del marciapiede i cinque livelli sono:

1. Ampio (consente il passaggio di 4 persone senza ostacoli)
2. Comodo (consente il passaggio di 3 persone con qualche ostacolo ma irrilevante)
3. Minimo (consente il passaggio di 2 persone ma con presenza di ostacoli che costringono ogni tanto a deviare il percorso)
4. Inadeguato (consente il passaggio di 1 persona con presenza di numerosi ostacoli lungo il percorso)
5. Assente (assente o in condizioni di impossibilità d'uso)

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

Tab. V Elenco degli attributi (variabili) di ogni segmento di strada.

Attributi (variabili)	Livelli (modalità)
X1 Larghezza utile del marciapiede	1 Ampio, 2 Comodo, 3 Minimo, 4 Inadeguato, 5 Assente
X2 Elementi di attrattività ambientale e urbana	1 Molti elementi, 2 Alcuni elementi, 3 Assenza di elementi, 4 Alcuni elementi di disturbo, 5 Molti elementi di disturbo
X3 Densità servizi, negozi, bar, attività commerciali	1 Abbondanti, 2 Alcuni, 3 Rari, 4 Assenti
X4 Opportunità di sosta (panchine, etc.)	1 Estesa, 2 Rada, 3 Assente
X5 Ripari e ombre	1 Forte, 2 Debole, 3 Assente
X6 Senso di marcia veicoli	1 Strada pedonale, 2 Senso unico, 3 Doppio senso
X7 Larghezza carreggiata veicoli	1 Strada pedonale, 2 Una corsia, 3 Due corsie, 4 Tre corsie, 5 Quattro corsie (o più)
X8 Limite di velocità	1 Strada pedonale, 2 ≤ 20 km/h, 3 30km/h, 4 50km/h, 5 ≥ 70 km/h.
X9 Corsia ciclabile (ciclabilità)	1 Corsia esclusiva su marciapiede, 2 Corsia su strada, 3 In strada condivisa con veicoli, 4 Non permesso
X10 Permeabilità pubblico-privato	1 Integrato, 2 Filtrato, 3 Separato
X11 Separazione pedoni-veicoli	1 Strada pedonale, 2 Forte, 3 Debole, 4 Assente
X12 Illuminazione	1 Eccellente, 2 Buona, 3 Inadeguata, 4 Assente
X13 Qualità pavimentazione marciapiede	1 Ottima, 2 Buona, 3 Mediocre, 4 Dissestata, 5 Assente
X14 Parcheggio su strada	1 Strada pedonale, 2 Permessso, 3 Proibito, 4 Illegale
X15 Riduttori di velocità fisica (dossi, attraversamenti rialzati, isole di traffico, miniroatorie)	1 Strada pedonale, 2 Abbondanti, 3 Alcuni, 4 Rari, 5 Assenti
X16 Riduttori di velocità non fisica (densità semafori, <i>enclosure</i>)	1 Strada pedonale, 2 Alta, 3 Media, 4 Bassa, 5 Assente
X17 Densità attraversamenti (opportunità di attraversamento)	1 Strada pedonale, 2 Forte, 3 Debole, 4 Assente
X18 Tipo di strada	1 Strada pedonale, 2 Strada carrabile

Per quanto riguarda la percezione soggettiva sintetica di qualità dello spazio pedonale, ai revisori è stato chiesto di esprimere il proprio giudizio valutativo complessivo su un gruppo di segmenti secondo le seguenti linee guida: «Per ogni via assegnata va espresso un giudizio sintetico di percezione della qualità e della camminabilità dal punto di vista del pedone. Il giudizio è espresso su una scala *qualitativa ordinale*, da 1 (insufficiente) a 5 (eccellente), tenendo conto delle *caratteristiche fisiche* del percorso pedonale; delle *caratteristiche qualitative* complessive dello spazio; ed in generale di quanto la strada è *sicura, comoda, piacevole, attrattiva, attraversabile*.

Per valutare la camminabilità *non* occorre considerare la distanza dal centro e gli elementi temporanei di disturbo (es. lavori in corso).»

Le risposte sono espresse su scala *Likert* a 5 punti, specificando il significato di ogni livello secondo le descrizioni guida:

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

5. Eccellente: massimo confort del pedone, strada molto piacevole da percorrere con elementi di particolare pregio e attrattività dello spazio urbano o del paesaggio circostante, nella quale è interessante sostare o passeggiare.
4. Molto Buona: strada confortevole percorrere, nella quale il transito è piacevole e privo di intralci.
3. Buona: strada percorribile, transito non troppo difficoltoso ma spazio non di particolare pregio.
2. Sufficiente: strada percorribile con difficoltà, ostacoli nel transito o qualità dello spazio scarsa.
1. Insufficiente: strada impossibile o molto insicura da percorrere, qualità dello spazio scarsa

4.1.1.1 Regressione lineare multipla

Per capire se esiste una relazione di dipendenza tra la qualità percepita dello spazio e le sue caratteristiche ci possiamo servire degli strumenti della statistica inferenziale, definendo la valutazione sintetica dei revisori come variabile dipendente e gli attributi dello spazio come variabili indipendenti.

Un primo test esplorativo è il calcolo del valore di correlazione tra il giudizio di percezione sulla qualità del LOS e le caratteristiche dello spazio. Diciassette dei diciotto indicatori hanno una correlazione positiva debole (0,30÷0,70) con il giudizio espresso dagli intervistati (Tabella VI); è perciò plausibile ipotizzare un rapporto tra le caratteristiche dello spazio con la qualità percepita, rapporto indagato più nel dettaglio adottando una regressione lineare multipla.

Tab. VI Matrice di correlazioni tra Variabile dipendente e Variabili indipendenti [X].

Var. [X]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
Correlaz.	0,63	0,56	0,46	0,40	0,40	0,43	0,38	0,51	0,57
Var. [X]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]
Correlaz.	0,31	0,60	0,44	0,54	0,45	0,50	0,22	0,56	0,45

Dopo aver normalizzato le variabili su scala da 0 a 1, è stata condotta una regressione lineare multipla⁵⁹ per esplorare la dipendenza tra la valutazione sintetica di qualità della camminabilità (variabile dipendente) e gli attributi dello spazio (variabile indipendente). Il primo modello, che impiega tutte le variabili indipendenti, produce un $R\text{-quadro}=0,60$ (modello A in Tabella VII). Un successivo modello è costruito con le 9 variabili significative del modello A ed ottiene un $R\text{-quadro}=0,59$ (modello B in Tabella VII). Questi risultati indicano che la percezione complessiva di camminabilità è fortemente associata alle seguenti nove variabili indipendenti: “X1 Larghezza utile del marciapiede”, “X2 Attrattività ambientale e urbana”, “X3 Densità servizi e attività”, “X4 Sosta (panchine)”, “X5 Ripari e ombra”, “X7 Larghezza carreggiata automobili”, “X9 Ciclabilità”, “X11 Separazione pedoni veicoli” e “X12 Illuminazione”.

Una rappresentazione grafica della relazione tra le caratteristiche dello spazio e i giudizi di qualità percepita è illustrata in Figura 4.2 dalle tabelle di contingenza a due vie tra ciascuna delle nove variabili e la variabile dipendente.

Per una comparazione delle relative importanze delle variabili indipendenti, sono state condotte delle regressioni singole per le 9 variabili significative. Da questo risultato, in Tabella VIII, si nota che può essere osservato un forte effetto singolo sulle variabili X1, X2, X3, X9 e X11, ognuna delle quali ottiene un $R\text{-quadro}>0,20$. Una regressione multipla per queste cinque variabili ha una bontà di adattamento $R\text{-quadro}=0,56$ (modello C in Tabella VIII). Questi risultati potrebbero indicare l'esistenza di una dipendenza tra le variabili, così come intuibile dalla matrice di correlazione in Tabella VI. Teniamo a mente questo aspetto, lo indagheremo più avanti, con delle procedure specifiche (paragrafo 4.1.2.4).

⁵⁹ Si parla di regressione singola quando una variabile dipendente è confrontata con una variabile indipendente. Si parla invece di regressione multipla quando una variabile dipendente è confrontata con più variabili indipendenti, mentre si ha una regressione multivariata quando più variabili dipendenti sono confrontate con un set di variabili indipendenti.

Tab. VII Risultati del modello di regressione lineare multipla.

	Modello A (R-quadro=0,60)				Modello B (R-quadro=0,59)				Modello C (R-quadro=0,56)			
	Stima	St. err	p-val.		Stima	St. err	p-val.		Stima	St. err	p-val.	
(Inter cetta)	-0,209	0,049	3E-05	***	-0,210	0,046	8E-06	***	-0,049	0,032	0,126	
X1	0,150	0,054	0,006	**	0,202	0,048	3E-05	***	0,277	0,046	3E-09	***
X2	0,229	0,059	1E-04	***	0,236	0,056	3E-05	***	0,287	0,056	4E-07	***
X3	0,085	0,033	0,011	*	0,101	0,032	0,002	**	0,130	0,032	6E-05	***
X4	0,096	0,047	0,042	*	0,107	0,046	0,020	*				
X5	0,066	0,037	0,077	,	0,067	0,034	0,048	*				
X6	-0,028	0,048	0,555									
X7	0,162	0,081	0,046	*	0,119	0,067	0,077	,				
X8	0,116	0,092	0,209									
X9	0,169	0,067	0,012	*	0,136	0,061	0,026	*	0,197	0,0572	6E-04	***
X10	0,001	0,034	0,968									
X11	0,141	0,069	0,042	*	0,129	0,055	0,019	*	0,174	0,048	4E-04	***
X12	0,169	0,055	0,002	**	0,212	0,050	3E-05	***				
X13	0,052	0,057	0,364									
X14	0,097	0,059	0,102									
X15	-0,048	0,071	0,498									
X16	-0,059	0,042	0,159									
X17	0,075	0,049	0,131									
X18	-0,094	0,066	0,160									

Se prestiamo attenzione ai nostri due obiettivi di ricerca (introduzione capitolo 4), sinora abbiamo trovato una risposta al primo quesito, abbiamo individuato quali sono le variabili significative nella valutazione della qualità dell'ambiente urbano. Ora, per rispondere al nostro secondo obiettivo – trovare un modello di previsione della qualità dello spazio – ci serviamo della possibilità di stimare i giudizi applicando la funzione ottenuta dalla regressione multipla.

La valutazione del LOS stimato è stata calcolata per ogni segmento stradale campionato nell'indagine, la Figura 4.3 mostra le differenze tra i giudizi dichiarati (y) e i giudizi stimati dal modello (\hat{y}).

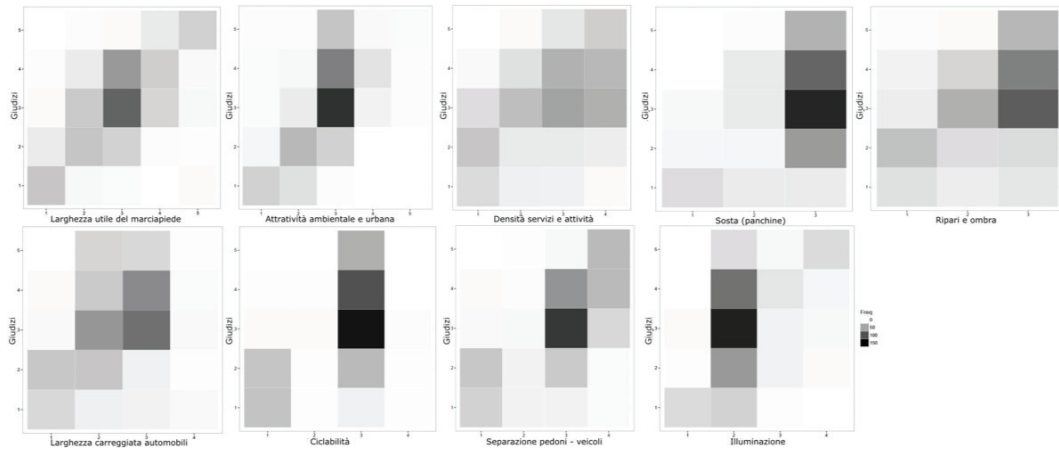


Fig. 4.2 Grafico a mattonelle della distribuzione di frequenze per giudizi e caratteristiche dello spazio.

Tab. VIII Regressione lineare monovariata per le variabili indipendenti più significative.

	X1	X2	X3	X4	X5	X7	X9	X11	X12
Coef. est.	0,6474	0,7637	0,3861	0,4598	0,3535	0,5540	0,6797	0,5804	0,5845
Std. error	0,0399	0,0559	0,0374	0,0520	0,0399	0,0676	0,0489	0,0379	0,0592
p-value	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	3,4E-15	<2e-16	<2e-16	<2e-16
(Intercetta)	0,2381	0,2017	0,3500	0,1979	0,3323	0,3509	0,1626	0,1939	0,2919
(Std- error)	0,0249	0,0317	0,0276	0,0483	0,0336	0,0340	0,0339	0,0290	0,0343
(p-value)	<2e-16	5,3E-10	<2e-16	5,1E-05	<2e-16	<2e-16	2,3E-06	7,3E-11	3,5E-16
R-quadro	0,3931	0,3152	0,2082	0,1617	0,1620	0,1418	0,3220	0,3659	0,1935

Come rilevato in figura non ci sono diversità significative, questo è confermato dalla differenza tra i giudizi reali e stimati, che presentano bassi valori e bassa dispersione (Tabella IX, modello B).

Graficamente, la frequenza dei residui è illustrata nell'istogramma in Figura 4.4; dai dati in figura, possiamo notare che il modello B predice approssimativamente il 55% dei segmenti come correttamente classificati (scarti: $-0,5 \div +0,5$), e classifica oltre il 95% dei segmenti al più in una classe oltre la valutazione sintetica assegnata dal valutatore (scarti: $-1,5 \div +1,5$) (Tabella IX).

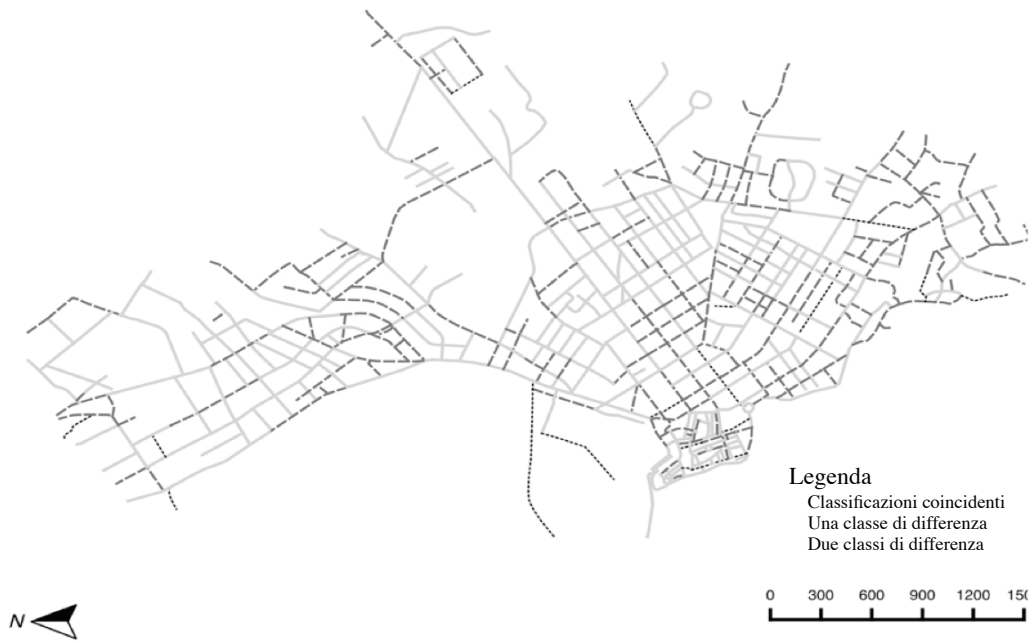


Fig. 4.3 Precisione del modello di predizione della camminabilità, risultati del modello B.

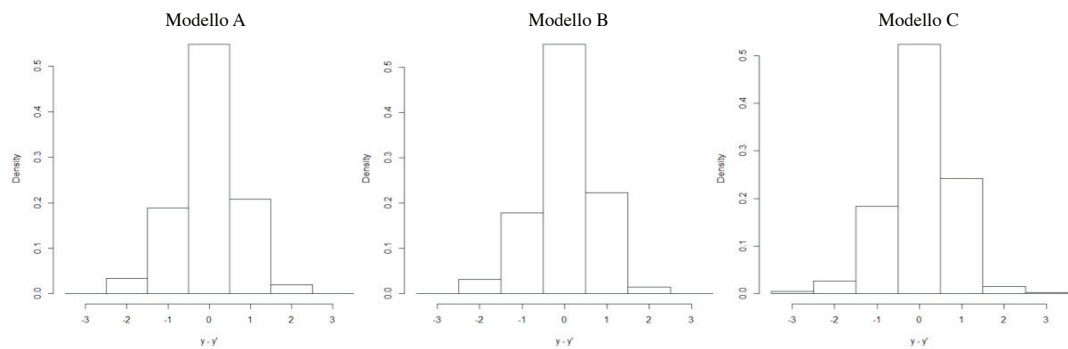


Fig. 4.4 Istogrammi dei residui regressione lineare multipla.

Tab. IX Residui del modello di regressione.

Residui	Mod. A	Mod. B	Mod. C	Classificazione	Mod. A	Mod. B	Mod. C
Min.	-2,26	-2,41	-2,96	Coincidente	54,90%	55,15%	52,45%
1°quartile	-0,44	-0,42	-0,49	1 classe oltre	39,71%	40,20%	42,66%
Mediana	-0,03	-0,07	-0,10	2 classi oltre	5,39%	4,65%	4,17%
3°quartile	0,48	0,50	0,56	3 classi oltre	0,00%	0,00%	0,74%
Max.	1,87	2,06	2,40				

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

4.1.2 Alcuni altri modelli di regressione e analisi

Per completezza, oltre il test di regressione lineare, possiamo sperimentare alcune altre procedure che vantano diverse proprietà d'interesse per il modello, in particolare vogliamo analizzare quattro aspetti:

- Studiare differenti procedure di regressione per verificare possibili differenze nella significatività delle variabili.
- Studiare differenti modelli di stima dei giudizi e verificare quale comporta i residui minori.
- Studiare la linearità delle variabili o la funzione che meglio approssima l'andamento dei dati.
- Studiare l'interazione tra le variabili per individuare falle nel modello di stima lineare (e compensativo) ottenuto dalla regressione multipla.

4.1.2.1 Regressione *stepwise*

Il primo metodo che ci interessa testare è la regressione *stepwise*. Basata sulla regressione multipla lineare permette di identificare il set di variabili che abbiano la migliore relazione con la variabile dipendente, cioè il maggior R-quadro possibile. Non ci sono particolari differenze con il set individuato dalla regressione lineare multipla, se non per l'assenza della variabile X4 e la presenza delle X8, X17 e X18; il valore di R-quadro ottenuto è 0,583 per il set di variabili: "X1 Larghezza marciapiede", "X2 Attrattività ambientale e urbana", "X3 Densità servizi e attività", "X5 Ripari e ombre", "X7 Larghezza carreggiata", "X8 Limite di velocità", "X9 Ciclabilità", "X11 Separazione pedoni-veicoli", "X12 Illuminazione", "X17 Densità attraversamenti" e "X18 Tipo di strada". La procedura restituisce i coefficienti in Tabella X.

Ciò che varia dalla regressione multipla è un numero maggiore di variabili (questo è dovuto al processo di *step* caratteristici della *stepwise*), che però non influenzano i risultati dell'analisi dei residui (Figura 4.5); da questo modello risulta che il 55,15% delle alternative è correttamente classificato, mentre il 95,83% è al più in una classe di errore, esattamente come per la regressione multipla. A parità di risultati è certamente preferibile un modello con meno variabili.

Tab. X. Risultati del modello di regressione *Stepwise*.

	Stima	St. err	t value	Pr(> t)	
(Int.)	-0,986	0,161	-6,009	0,000	***
X1	0,164	0,047	3,469	0,001	**
X2	0,221	0,055	4,002	0,000	***
X3	0,082	0,032	2,555	0,011	.
X5	0,076	0,035	2,182	0,030	.
X7	0,125	0,055	2,269	0,024	.
X8	0,287	0,078	3,693	0,000	***
X9	0,209	0,061	3,429	0,001	**
X11	0,242	0,064	3,773	0,000	***
X12	0,189	0,050	3,759	0,000	***
X17	0,121	0,046	2,613	0,009	*
X18	0,751	0,143	5,258	0,000	***

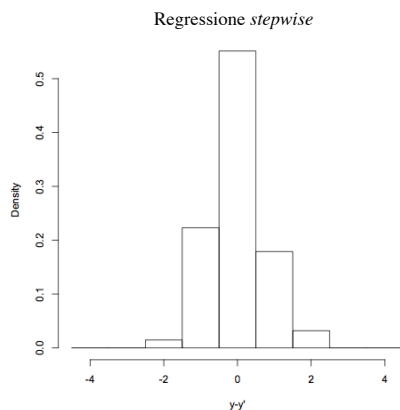


Fig. 4.5 Istogramma dei residui regressione *stepwise*.

4.1.2.2 Regressione logistica ordinale

La terza metodologia di regressione testata è la regressione logistica ordinale (*ordered logistic regression*). Questa permette di definire la relazione tra le variabili indipendenti ed una variabile dipendente ordinale (discreta) politomica. Sappiamo che una caratteristica delle variabili ordinali è che la differenza tra ogni modalità non è necessariamente la stessa, proprio come può essere una categoria di risposta basata sulle percezioni. Quando eseguiamo una regressione lineare con

variabili ordinali stiamo supponendo che la differenza di importanza tra le classi sia costante, il che non è sempre valido (non è infatti una scala cardinale). Per superare questo limite, la regressione ordinale consente – a partire da un modello logistico o probabilistico (*logit* o *probit*) – di incorporare la natura ordinale della variabile dipendente attraverso l'analisi dei rapporti di probabilità (*odds*), cioè di osservare un giudizio, o uno inferiore, rispetto ai giudizi superiori⁶⁰, formalmente: $\theta_j = \text{prob}(\text{score} \leq j) / \text{prob}(\text{score} > j)$, dove j indica il valore della variabile dipendente.

Il metodo di analisi della regressione ordinale non prevede la stima di un valore di R-quadro che potremmo confrontare con la regressione lineare multipla, esistono alcuni pseudo R-quadro che possono essere analizzati, ma i loro risultati non sono sempre coincidenti ed affidabili (Norusis, 2011, p.81) per un confronto con i risultati di una regressione basata sulla stima dei minimi quadrati OLS⁶¹. Ad ogni modo sono riportati per avere un'indicazione approssimativa che, insieme agli altri dati sulla bontà di adattamento del Chi-quadro, possono aiutarci a comprendere i risultati.

Tenendo a mente che lo scopo principe è quello di definire una formula di stima della qualità dei segmenti stradali a partire dalle sue caratteristiche, un primo modello è stato costruito con tutte le 18 variabili ed un secondo con le 9 variabili significative del modello lineare B. Un confronto tra gli scarti ci può dare qualche indicazione sulla miglior stima dei risultati.

Il primo modello A (Tabella XI, Figura 4.6), vanta dei pseudo R-quadro di 0,743 (Nagelkerke), 0,706 (Cox e Snell) e 0,408 (McFadden). Valori tutto sommato rassicuranti supportati dalla significatività del test del Chi-quadro = 0,017⁶² che però garantisce la significatività (la differenza tra le distribuzioni, teorica e

⁶⁰ Ad esempio date tre classi $A > B > C$ gli *odds* saranno calcolati da:

$$\theta_C = \text{prob}(C) / \text{prob}(A \wedge B)$$

$$\theta_B = \text{prob}(C \vee B) / \text{prob}(A)$$

$$\theta_A = \text{prob}(C \vee B \vee A) = 1$$

⁶¹ OLS: *Ordinary Least Squares*.

La regressione ordinale è invece basata sulla massima verosimiglianza (*Maximum likelihood*).

⁶² I valori del test del Chi-quadro possono non essere affidabili a causa della dipendenza tra alcune variabili. La validità dell'andamento del modello è incerta.

osservata, non è casuale) solo al 5%. Nonostante i risultati è comunque possibile stimare i giudizi in funzione dei coefficienti registrati; gli scarti tra giudizi rilevati e predetti restituiscono il 61,5% delle alternative come correttamente classificate ed il 98,0% al più in una classe di errore.

Con un secondo modello B (Tabella XI, Figura 4.6), ottenuto dall'analisi delle 9 variabili significative della regressione lineare, abbiamo dei valori di pseudo R-quadro di 0,675, 0,641 e 0,342 (risp. Nagelkerke, Cox e Snell, McFadden) e una significatività del test Chi-quadro = 0,000. In questo caso il modello guadagna di attendibilità statistica ma perde precisione nella classificazione, che si attesta su 58% di alternative correttamente classificate e 96% di alternative al più in una classe di errore.

Si sarà intuito che la regressione logistica analizza variabili *dummy*, ovvero variabili binarie ottenute scomponendo le variabili iniziali. Per essere più chiari, la variabile primitiva “X1 Larghezza del marciapiede” è composta da cinque modalità (assente, inadeguato, minimo, comodo, ampio), nella regressione ordinale avremo perciò le cinque variabili dicotomiche: “X1.5 Marciapiede assente”, “X1.4 Marciapiede inadeguato”, “X1.3 Marciapiede minimo”, “X1.2 Marciapiede comodo”, “X1.1 Marciapiede ampio”, tutte con modalità Vero, Falso (ovvero Sì, No). Per questo motivo – per un test esplorativo – è possibile escludere quelle variabili indipendenti (ovvero le modalità delle variabili originarie) che non risultano significative⁶³. Significa analizzare solo una parte delle caratteristiche dello spazio che risultano realmente coinvolte nella definizione del giudizio di qualità, in maniera ancora più dettagliata rispetto alle procedure precedenti⁶⁴. Un modello C (Tabella XI, Figura 4.6) che segue questa

⁶³ Oltre un possibile incremento della qualità del modello e una riduzione di probabili errori, ridurre il numero di caratteristiche dello spazio analizzate snellisce le procedure di *audit*.

⁶⁴ Una critica che possiamo muovere a questa procedura è che, con i dati del nostro campione, rischiamo di perdere alcuni elementi dello spazio che sono importanti in altre circostanze. Se però immaginassimo di applicare la procedura in maniera dinamica (e.s. *machine learning*) riusciremmo a cogliere ogni aspetto nel migliore dei modi.

regola include le 12 variabili⁶⁵: “X1.1 Marciapiede ampio”, “X1.2 Marciapiede comodo”, “X1.3 Marciapiede minimo”, “X3.1 Densità di servizi abbondante”, “X5.1 Ripari e ombra forte”, “X5.2 Ripari e ombra debole”, “X7.1 Strada pedonale”, “X7.2 Strada una corsia”, “X7.3 Strada due corsie”, “X11.3 Separazione pedoni-veicoli debole”, “X12.1 Illuminazione eccellente” e “X12.2 Illuminazione buona”, per cui pseudo R-quadro è 0,608 (Cox e Snell), 0,577 (Nagelkerke) e 0,288 (McFadden). Il Chi-quadro di 2682,592 per 432 gradi di libertà ottiene un $P\text{-value} = 0,000$ che ci permette di dire che esiste una certa dipendenza tra le variabili. Il modello predice il 54% delle alternative nella classe corretta e il 96% al più in una classe superiore o inferiore rispetto ai giudizi assegnati dai valutatori.

Abbiamo quindi costruito un modello che analizza solo le variabili *dummy* significative, ma lo abbiamo fatto in maniera parziale ed esplorativa. Non è utile proseguire in questo test visti i migliori risultati in altri modelli di previsione, anche se per completezza sarebbe opportuno definire il modello senza intercetta, che è però possibile solo con variabili continue. Questa applicazione richiederebbe quindi un intervento sulle variabili che non è possibile in questo contesto.

Per evitare possibili interpretazioni errate, è importante sottolineare che non è possibile confrontare la significatività degli attributi della regressione ordinale con i modelli precedentemente proposti, anche se questa sarebbe un’operazione auspicabile. La stima dei parametri di una regressione logistica – in maniera quasi analoga alla regressione lineare – restituisce i valori dei coefficienti di ogni variabile (in questo caso *dummy*) ed il suo valore di significatività (test di Wald). Ma a differenza della regressione lineare dove β rappresenta la variazione media di Y al crescere di un’unità di X , nella regressione logistica, β esprime la variazione del *logit* corrispondente ad un incremento unitario di X , ovvero misura la variazione nel *logit* dovuta al possesso dell’attributo descritto da X . Questo si riflette nell’impossibilità, senza manipolare i coefficienti, di fare confronti tra variabili, ma permette solo di confrontare le modalità (di una variabile) tra loro.

⁶⁵ A causa della così detta “trappola della variabile *dummy*” il modello non include l’ultima modalità come variabile. Il suo valore è integrato nell’intercetta quando tutte le variabili *dummy* di una stessa variabile originaria sono zero. Un modo alternativo per specificare la variabile è di costruire un modello senza intercetta ma questo è possibile quando le variabili sono continue (si parla di covariate e non di fattori).

Tab. XI Risultati della regressione ordinale.

Modello A				Modello A (continua)			
	Stima	Errore std	sing		Stima	Errore std	sing
[G = 1,00]	-11,059	2,062	0,000 ***	[X9=2,00]	3,984	1,654	0,016 *
[G = 2,00]	-7,783	2,041	0,000 ***	[X9=3,00]	2,763	1,360	0,042 *
[G = 3,00]	-4,622	2,026	0,023 *	[X10=1,00]	0,164	0,376	0,663
[G = 4,00]	-1,892	1,999	0,344	[X10=2,00]	-0,152	0,368	0,680
[X1=1,00]	-1,890	0,931	0,042 *	[X11=1,00]	-8,709	2353	0,997
[X1=2,00]	0,235	0,824	0,776	[X11=2,00]	-0,468	0,631	0,459
[X1=3,00]	0,383	0,792	0,629	[X11=3,00]	-0,986	0,411	0,016 *
[X1=4,00]	1,258	0,766	0,101	[X12=1,00]	-4,505	1,031	0,000 ***
[X2=1,00]	-1,594	1,052	0,130	[X12=2,00]	-1,134	0,616	0,066 .
[X2=2,00]	-0,665	1,021	0,515	[X12=3,00]	-0,313	0,632	0,621
[X2=3,00]	0,264	0,982	0,788	[X13=1,00]	-4,698	1,243	0,000 ***
[X2=4,00]	0,768	0,981	0,434	[X13=2,00]	-3,410	1,084	0,002 **
[X3=1,00]	-1,430	0,438	0,001 **	[X13=3,00]	-3,415	1,083	0,002 **
[X3=2,00]	-0,301	0,347	0,385	[X13=4,00]	-4,446	0,997	0,000 ***
[X3=3,00]	0,061	0,297	0,838	[X14=1,00]	-2,933	1,216	0,016 *
[X4=1,00]	0,363	0,686	0,597	[X14=2,00]	-0,478	0,489	0,328
[X4=2,00]	0,450	0,375	0,230	[X14=3,00]	-0,150	0,545	0,784
[X5=1,00]	-1,141	0,490	0,020 *	[X15=2,00]	-1,321	0,929	0,155
[X5=2,00]	-0,740	0,294	0,012 *	[X15=3,00]	0,875	0,988	0,376
[X6=1,00]	-1,175	0,906	0,194	[X15=4,00]	0,076	0,764	0,921
[X6=2,00]	0,407	0,296	0,169	[X16=1,00]	1,895	1,534	0,217
[X7=1,00]	20,112	2353	0,993	[X16=2,00]	0,299	0,420	0,477
[X7=2,00]	-1,994	0,752	0,008 **	[X16=3,00]	-0,285	0,387	0,462
[X7=3,00]	-1,808	0,720	0,012 *	[X16=4,00]	-0,101	0,417	0,808
[X8=2,00]	-1,646	0,990	0,096 .	[X17=1,00]	-11,040	0,000	.
[X8=3,00]	-0,404	0,539	0,453	[X17=2,00]	-1,298	0,398	0,001 **
[X9=1,00]	-0,575	1,114	0,606	[X17=3,00]	-0,754	0,361	0,037 *

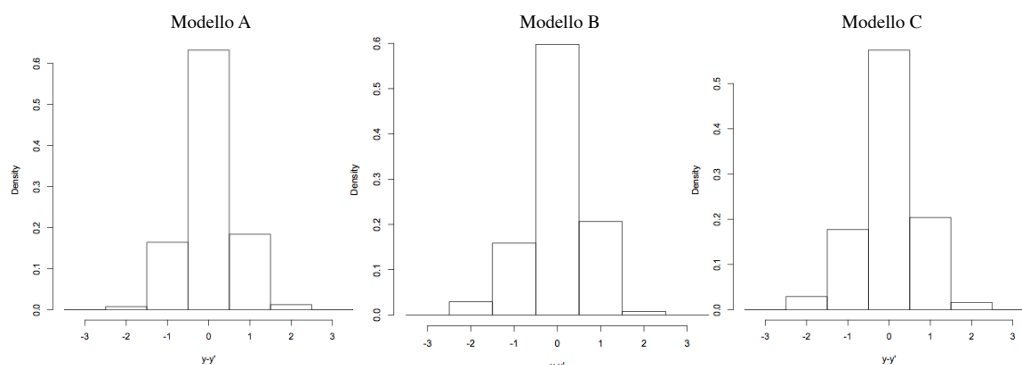


Fig. 4.6 Istogrammi dei residui regressione ordinale.

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
 Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

Tab. XI (continua)

		Modello B			Modello C				
		Stima	Errore std	Sign.		Stima	Errore std	Sign.	
Soglia	[G = 1,00]	14,572	4,2	0,001 ***		16,175	1,986	0 ***	
	[G = 2,00]	17,406	4,21	0 ***		18,534	2,035	0 ***	
	[G = 3,00]	20,243	4,235	0 ***		21,126	2,095	0 ***	
	[G = 4,00]	22,654	4,273	0 ***		23,365	2,147	0 ***	
Ubicazione	[X1.1=,00]	2,876	0,728	0 ***		3,852	0,483	0 ***	
	[X2.1=,00]	1,622	0,843	0,054 .					
	[X3.1=,00]	1,277	0,386	0,001 ***		1,587	0,307	0 ***	
	[X4.1=,00]	0,077	0,621	0,901					
	[X5.1=,00]	1,145	0,405	0,005 **		1,251	0,39	0,001 ***	
	[X7.1=,00]	-8,199	1,002	0 ***		3,315	0,679	0 ***	
	[X9.1=,00]	1,284	1,076	0,233					
	[X11.1=,00]	9,619	0						
	[X12.1=,00]	5,122	0,998	0 ***		6,093	0,872	0 ***	
	[X1.2=,00]	1,58	0,604	0,009 **		1,965	0,337	0 ***	
	[X2.2=,00]	0,719	0,813	0,376					
	[X3.2=,00]	0,445	0,314	0,156					
	[X4.2=,00]	-0,115	0,341	0,735					
	[X5.2=,00]	0,702	0,273	0,01 **		0,532	0,249	0,033 *	
	[X7.2=,00]	2,062	0,652	0,002 **		2,099	0,621	0,001 ***	
	[X9.2=,00]	-1,696	1,499	0,258					
	[X11.2=,00]	0,869	0,554	0,117					
	[X12.2=,00]	1,325	0,54	0,014 *		1,004	0,277	0 ***	
	[X1.3=,00]	1,434	0,562	0,011 *		1,373	0,272	0 ***	
	[X2.3=,00]	-0,363	0,747	0,627					
	[X3.3=,00]	-0,004	0,27	0,99					
	[X7.3=,00]	1,684	0,644	0,009 **		1,574	0,611	0,01 **	
	[X9.3=,00]	-1,583	1,275	0,214					
	[X11.3=,00]	1,176	0,298	0 ***		0,766	0,238	0,001 ***	
[X12.3=,00]	0,589	0,551	0,285						
[X1.4=,00]	0,388	0,558	0,487						
[X2.4=,00]	-0,453	0,801	0,571						

Nota: Le variabili per cui i fattori sono ridondanti sono escluse dalla tabella

4.1.2.3 Test di modelli non lineari

Abbiamo ragione di credere che la relazione tra le variabili indipendenti e quella dipendente possa non essere lineare (ne abbiamo un'idea visiva dai grafici a mattonelle, Figura 4.2), è perciò utile testare questo aspetto. Dobbiamo però essere consapevoli di effettuare un test “pericoloso” da interpretare cautamente. Per applicare questa procedura è necessario considerare l'assunto che la variabile dipendente sia scalare, sappiamo che potrebbe non essere così, ma possiamo supporre che la differenza possa non essere tanto marcata; è certamente un punto sul quale indagare ancora per test più approfonditi. In particolare, possiamo effettuare un test con la procedura di regressione della “curva stimata”, che restituisce le statistiche di regressione per diversi modelli non lineari. Questo ci permette – con alcune precauzioni d'interpretazione⁶⁶ – di avere un'idea più chiara delle possibili relazioni. Per questo test ci limitiamo a valutare le nove variabili significative della regressione lineare, poiché sono quelle che ci interessano per la costruzione del modello.

Consci della difficoltà di fare stime con elevata precisione, è bene limitarsi all'analisi di alcune forme di funzione generiche, la funzione logaritmica e la funzione potenziale. È infatti sensato sostenere che il giudizio marginale (variabile dipendente) possa essere crescente o decrescente e non costante. Ad esempio, la funzione potrebbe essere potenziale se all'aumentare del valore della variabile il giudizio aumentasse più che proporzionalmente, oppure logaritmica se il giudizio aumentasse in maniera sempre minore. Significherebbe che la relazione non è lineare e che una modalità influenza in maniera maggiore il giudizio rispetto alle altre.

Tab. XII Valori di R-quadro per i test di regressione non lineari.

	X1	X2	X3	X4	X5	X7	X9	X11	X12
Lineare	0,393	0,315	0,208	0,161	0,162	0,142	0,322	0,366	0,193
Logaritmica	0,407	0,298	0,218	0,166	0,160	0,182	0,333	0,339	0,228
Potenza	0,425	0,370	0,209	0,219	0,152	0,196	0,399	0,348	0,235

⁶⁶ Le procedure di regressione testate nascono come applicazioni per l'analisi con variabili continue, in questo caso sono applicate a variabili discrete.

Dati i valori di R-quadro per ogni funzione (Tabella XII), notiamo che la maggioranza delle relazioni è spiegata con miglior precisione da una funzione potenziale, per cui all'aumentare della variabile x l'incremento marginale della variabile y decresce. Intendendo questi dati per variabili discrete (ordinali) possiamo interpretare come meno influenti le modalità inferiori e più decisive quelle superiori⁶⁷; un esempio può aiutare: per la variabile "X5 opportunità di sosta" non ci sono grosse differenze quando le sedute sono "assenti" o "rade" (la differenza di giudizio è perciò influenzata da un altro fattore), mentre si nota un forte incremento del giudizio quando la *sedibilità* è "estesa" (si veda grafico a mattonelle). In questo caso significa che la presenza di molte panchine ha un'importante rilevanza nella decisione⁶⁸.

La necessità di prendere con estrema cautela questi risultati è chiara, stiamo infatti analizzando delle procedure per variabili continue su misure discrete, i risultati sono perciò molto approssimativi, ma possono comunque essere utili per comprendere l'andamento generale dei dati. Riflettere sulla linearità dei dati ci spinge a concentrare la ricerca su metodi che analizzano le variabili dicotomiche (regressione ordinale) o metodi non-compensativi (si veda paragrafo 4.1.3) e quindi non basati sulla linearità delle relazioni.

4.1.2.4 Dipendenza tra le variabili

Oltre alla linearità della funzione, ci interessa capire se esiste un'interazione tra le variabili indipendenti e se quindi i risultati possono essere influenzati da

⁶⁷ Se le variabili fossero continue e non discrete significherebbe che giudizi "minori" sono molto simili, distinguendosi solo con incrementi notevoli della variabile x , mentre giudizi "maggiori" si otterrebbero con un ridotto incremento della x .

⁶⁸ Secondo la tabella delle frequenze il giudizio per le modalità 1 e 2 è molto simile, mentre cambia radicalmente per la modalità 3.

	X5=1	X5=2	X5=3	
G=5	0%	0%	10%	11%
G=4	0%	3%	22%	25%
G=3	1%	3%	31%	36%
G=2	2%	2%	14%	17%
G=1	4%	3%	3%	11%
	8%	12%	80%	

multicollinearità⁶⁹. Un test di correlazione tra le variabili può spiegarci qualcosa in più e un modello più complesso, basato sulle equazioni lineari (*linear equation model*), può essere utile per testare se un certo diagramma dei percorsi (*path diagram*) possa essere significativo e rappresentare le relazioni tra le variabili indipendenti e dipendenti⁷⁰.

Il test di correlazione tra le variabili può aiutarci a capire se alcune combinazioni tra modalità di attributi siano più frequenti di altre e se quindi possa esserci un legame tra due variabili. Testando la correlazione delle nove variabili significative con il test Tau-b di Kendall⁷¹ si ottengono i risultati in Tabella XIII.

Tab. XIII Matrice di correlazione tra le variabili.

	X1	X2	X3	X4	X5	X7	X9	X11	X12
X1	1	,442**	,369**	,301**	,275**	,246**	,439**	,503**	,352**
X2		1	,357**	,364**	,274**	,273**	,475**	,405**	,284**
X3			1	,236**	,258**	,134**	,203**	,307**	,273**
X4				1	,150**	-0,056	,343**	,154**	,303**
X5					1	,317**	,343**	,409**	,125**
X7						1	,430**	,450**	-0,006
X9							1	,515**	,219**
X11								1	,298**
X12									1

** La correlazione è significativa a livello 0,01 (a due code)

È evidente che esista una relazione, seppur debole, tra le diverse caratteristiche dello spazio; non sorprende! L'esperimento condotto sul campo e non su

⁶⁹ Il problema della multicollinearità riguarda modelli statistici espressi mediante un'equazione lineare, quando alcune o tutte le variabili sono fra loro fortemente correlate. Le conseguenze sono: influenze nella significatività dei singoli coefficienti che risultano modesti e nel *fitting* della regressione che risulta elevato (si osservano elevati valori dell'indice R²).

⁷⁰ Per approfondimenti si veda: Loehlin, J. (2004) *Latent variable models, an introduction to Factor, Path and Structural Equation Analysis*, Lawrence Erlbaum Associates Publisher, London.

⁷¹ È preferibile rispetto al test di Pearson quando i dati non sono normalmente distribuiti o prevedono categorie ordinate.

alternative ortogonali⁷² non permette di escludere quelle influenze sulla variabilità campionaria derivante dalle caratteristiche della città. Possiamo dire che ci sono (almeno) due motivi per cui si ottengono questi risultati:

- Diversi tipi di fenomeni urbani (sociali, di gestione, ecc.) compongono lo spazio e concentrano le caratteristiche in diversi punti della città, che diventano luoghi di aggregazione, che si abbandonano, che sono monofunzionali, ecc. Ad esempio, nel centro città c'è solitamente una tendenza di aggregazione degli elementi di attrattività (centri commerciali naturali) e/o c'è un'attenzione maggiore nella cura dello spazio (i "salotti urbani"). Questo porta a una maggior frequentazione dei luoghi e di conseguenza qualità superiore. Le periferie sono invece abbandonate, dimenticate, senza servizi, poco frequentate e quindi di scarsa qualità.

- Più generalmente, alcune caratteristiche dello spazio sono spesso complementari, o si escludono a vicenda; una bassa qualità dello spazio si accompagna spesso ad una ridotta manutenzione della strada, l'assenza del marciapiede ad un'alta velocità dei veicoli, un elemento di pregio ambientale alla presenza di sedute per ammirarlo. Questo ci spinge ad approfondire lo studio delle relazioni e tentare di stimare un possibile diagramma delle interazioni tra le variabili.

4.1.2.5 Relazione tra le variabili, il *path diagram*

Per indagare la relazione tra le diverse variabili può essere adottato un modello esplorativo ad equazioni strutturali SEM (*structural equation modeling*), in particolare la *path analysis*. Questo strumento permette di studiare l'intensità delle relazioni e le dipendenze tra le variabili del sistema, oltre che definire la bontà del modello esaminato; i suoi dettagli metodologici sono presentati in appendice A. La *path analysis* ha il vantaggio di esaminare sia gli effetti di relazione diretti che quelli indiretti, questi ultimi sono presenti quando una variabile ha effetti su altri caratteri attraverso l'influenza di variabili intermedie. Con questa procedura possiamo testare la relazione tra le variabili che hanno ottenuto forti effetti di

⁷² Per ortogonale si intende un disegno dell'esperimento per cui ogni modalità si ripete un numero di volte uguali ed in maniera alternata rispetto alle modalità delle altre variabili.

relazione nella regressione singola e che sono risultate correlate tra loro, ovvero le variabili del modello C⁷³.

Nella nostra ipotesi, nel modello finale ottenuto applicando le classiche procedure di miglioramento del modello della analisi LISREL (*Linear Structural RELationship*), la variabile “giudizio” è direttamente influenzata dalle cinque variabili considerate nel modello, la X1, X2, X3, X9, X11. L’“attrattività ambientale e urbana” è considerata esogena (c’è ragionevole certezza nel dire che la *qualità ambientale* non è influenzata dagli altri caratteri) ed ha una diretta influenza su tutte le altre variabili, che sono endogene. La “larghezza del marciapiede” influenza la “densità dei servizi”, la “separazione tra pedoni e veicoli” e la “ciclabilità”. Quest’ultima è a sua volta influenzata dalla “separazione tra pedoni e veicoli” e dalla “densità dei servizi”. Le relazioni sono rappresentate graficamente in Figura 4.7 e descritte di seguito in forma di equazioni strutturali.

$$G = p_{g1} * X1 + p_{g3} * X3 + p_{g9} * X9(11) + p_{g11} * X11(14) + p_{g2} * X2 + e_g$$

$$X1 = p_{12} * X2 + e_1$$

$$X3 = p_{13} * X1 + p_{23} * X2 + e_3$$

$$X9(11) = p_{19} * X1 + p_{39} * X3 + p_{911} * X11(14) + p_{29} * X2, + e_{11}$$

$$X11(14) = p_{g11} * X1 + p_{g211} * X2, + e_{14}$$

Nell’equazione strutturale gli effetti causali diretti sono rappresentati dal coefficiente p , altrimenti chiamato *path coefficient* o *structural coefficient*⁷⁴; può

⁷³ È bene limitarsi alle cinque variabili del modello C e non estendendo il test alle nove variabili del modello B poiché un elevato numero di interazioni tra le variabili riduce la precisione del modello che può risultare significativo anche quando non lo è.

⁷⁴ Il principio che regola questo metodo è quello di esprimere le covarianze o le correlazioni tra due variabili come somma di tutti i percorsi composti che legano le due variabili, tramite dei coefficienti collocati sui suddetti percorsi, detti *path coefficients*.

I *path coefficients* esprimono la portata del nesso causale o della pura associazione tra variabili. Rappresentano, per ogni equazione presa in considerazione, dei veri e propri coefficienti di regressione che consentono di valutare l’impatto di un qualsiasi cambiamento di una variabile esogena sulla corrispondente endogena; sono dei coefficienti di regressione parziale, essendo la loro somma pari alla correlazione totale. Pertanto, essi valutano l’impatto di una variabile su un’altra quando tutte le altre, nel sistema, sono mantenute costanti.

essere considerato analogo al coefficiente di regressione standardizzato (beta) di una regressione multipla, consente quindi di valutare l'impatto di un qualsiasi cambiamento di una variabile esogena sulla corrispondente endogena. Il valore di un *path coefficient* indiretto è determinato dal prodotto dei coefficienti del percorso. Ad esempio, in Figura 4.7, i *paths* da X2 a X1 (p_{12}) e da X1 a X3 (p_{13}) combinati insieme producono l'effetto indiretto di X2 su X3 (uguale a $p_{12}p_{13} = 0,544 \cdot 0,324 = 0,18$). Infine il valore residuo di errore (*disturbance term*) rappresenta la componente stocastica non inclusa nel modello.

Nel *path diagram* in Figura 4.7, per ogni relazione tra variabili, sono riportati i valori dei *paths coefficients*, dei coefficienti standardizzati (*standardized coefficients*) tra parentesi quadre e il valore di significatività (*t-value*) tra parentesi graffe.

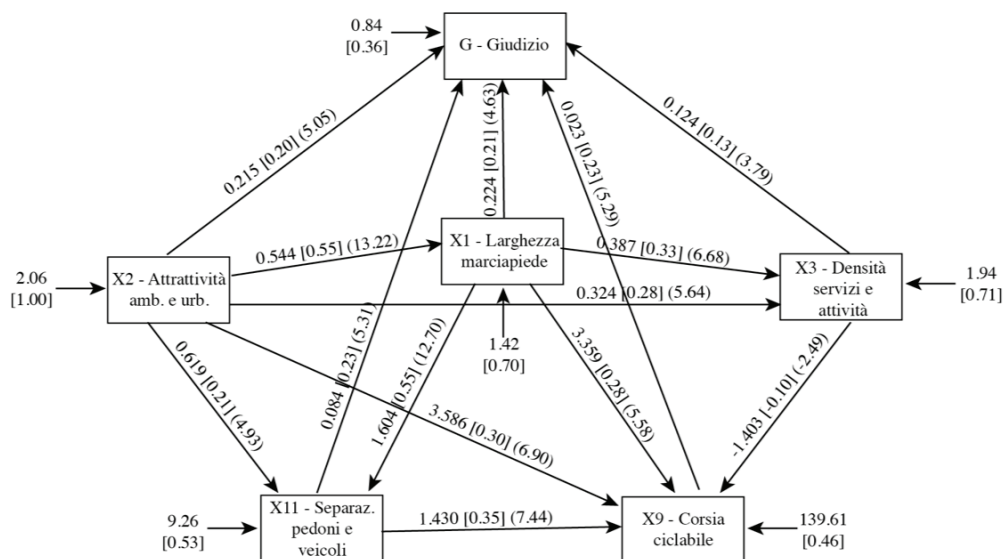


Fig. 4.7 *Path diagram* per le variabili del modello C.

A seguire sono richiamate le cinque equazioni strutturali per ogni variabile endogena, nel quale il valore del coefficiente p è sostituito con il valore ottenuto dal calcolo. Gli *Errorvar* indicano gli errori stocastici di ogni variabile mentre R^2 è il valore della bontà di adattamento per l'equazione.

$$G = 0,224*X1 + 0,124*X3 + 0,0203*X11 + 0,0843*X14 + 0,215*X2,$$

$$Errorvar. = 0,837, R^2 = 0,641$$

$$X1 = 0,544*X2, Errorvar. = 1,423, R^2 = 0,300$$

$$X3 = 0,387*X1 + 0,324*X2, Errorvar. = 1,942, R^2 = 0,292$$

$$X11 = 3,259*X1 - 1,043*X3 + 1,430*X14 + 3,586*X2, Errorvar. = 139,614, R^2 = 0,535$$

$$X14 = 1,604*X1 + 0,619*X2, Errorvar. = 9,261, R^2 = 0,471$$

I risultati significativi del modello ci portano ad individuare come effetti principali diretti sulla variabile dei giudizi G la larghezza del marciapiede X1 e l'attrattività ambientale e urbana X2 (in accordo con i coefficienti β della regressione multipla). I maggiori effetti di relazione diretti sono tra la variabile X2 che influenza la variabile X1 [0,55] e tra la variabile X1 che ha un effetto sulla variabile X11 [0,55]. Altri effetti considerevoli sono quelli dell'attrattività dello spazio [0,30] e della separazione tra pedoni e veicoli [0,35] nei confronti della ciclabilità, oltre che l'effetto della larghezza del marciapiede sulla densità di negozi e attività [0,33]. L'unica relazione indiretta è tra la densità dei servizi che influisce negativamente [-0,10] sulla ciclabilità (nelle zone commerciali si tende a non costruire piste ciclabili). Le relazioni indirette più significative con il giudizio sono quelle da X2 passante per X1 [0,12], X11 passante per X9 [0,08] e X1 passante per X11 [0,13]. La variabile X1 si dimostra perciò una delle variabili più significative nell'analisi di importanza, ha un forte effetto diretto con la variabile giudizio ed ha forti influenze verso le altre variabili endogene. La variabile X2 è forse la variabile più influente all'interno del modello, ha infatti forti effetti su tutte le variabili endogene. La variabile X3 è mediamente influente sul valore di giudizio e gioca un ruolo medio per quanto riguarda gli effetti indiretti delle altre due variabili sul giudizio di qualità dello spazio. La variabile X9 risulta la meno influente direttamente, ma ha un forte effetto indiretto dalla variabile X1 alla variabile G. Infine la variabile X11 ha un ridotto effetto diretto ma è la variabile più influente negli effetti indiretti di G da X1 e X2.

Per testare la validità dei percorsi proposti, questo tipo di analisi si basa sul test del χ^2 (della massima verosimiglianza tra la matrice di covarianza/correlazione⁷⁵ dei dati e quella teorica ottenuta dal modello), ma a causa dei problemi di sensitività del test⁷⁶, la misura più popolarmente utilizzata è l'indice di adattabilità *Root Mean Square Error of Approximation* $RMSEA = \sqrt{(d/df)}$. Dove d è ottenuto dividendo $(\chi^2 - df)$ per $(N - 1)$.

Quindi, nel nostro caso con un $\chi^2=3,16$ ottenuto con 1 df per un campione di 408 unità, il parametro $d = (3,16 - 1)/408 = 0,005294$ per cui $RMSEA = \sqrt{(0,005/1)} = 0,0728$. Per interpretare il valore di $RMSEA$, i valori $\leq 0,05$ indicano comunemente un *fit* del modello molto buono, i valori $\leq 0,10$ e $\geq 0,05$ un *fit* buono (Loehlin 2004, p.69).

Questa analisi ha portato ad un modello con valore Chi-quadro = 3,16, a 1 grado di libertà ($< 3,84$, $P = 0,05$) con un P -value = 0,0753 e $RMSEA = 0,0728$.

Sia il valore di Chi-quadro che il valore di $RMSEA$ indicano il modello come significativo, accettando l'ipotesi H_0 che descrive il modello. Significa quindi che le variabili si influenzano a vicenda e che un modello di previsione compensativo potrebbe non essere un ottimo strumento di misura. L'alternativa è un modello non-compensativo, alcuni sono descritti e testati a partire dal sottoparagrafo seguente.

4.1.3 Compensazione

Ciò che possiamo dedurre dalle analisi condotte sino ad ora è che una procedura lineare soffre di alcuni limiti di interazione tra le variabili e di non linearità delle relazioni. Oltre questo, le procedure che abbiamo applicato sono compensative e, sotto alcuni punti di vista del nostro obiettivo, questo può essere ritenuto un limite; nella pratica urbanistica potrebbe non essere un requisito desiderabile, ma sarebbe piuttosto preferito un metodo non-compensativo. Il termine compensativo

⁷⁵ In questa analisi è stata usata la correlazione e non la covarianza perché è il metodo preferito per ignorare la varianza delle variabili esaminate.

⁷⁶ "With large enough samples, substantively trivial discrepancies can lead to rejection of an otherwise highly satisfactory model; with small enough samples, χ^2 can be non significant even in the face of gross misfits" (Loehlin 2004, p.60).

sta ad indicare – nella teoria della misura – un metodo che compensa cattivi risultati con ottimi risultati e viceversa, definendo come buona sia un’alternativa con tutte le caratteristiche buone sia un’alternativa con alcune caratteristiche pessime ed altre ottime. Un metodo non-compensativo è invece un metodo che non media i valori di ogni attributo, considerando in un certo modo la dispersione tra i valori di ogni variabile. Un metodo non-compensativo non permette di classificare un’alternativa oltre un certo livello se questa presenta determinate caratteristiche che sono – per il decisore – insufficienti per raggiungere il livello; in generale – un metodo non-compensativo – classifica le alternative secondo regole di assegnazione che non prevedono la somma tra i valori di ogni attributo. Ma vediamo con un esempio il perché un metodo non-compensativo sarebbe da preferire ad uno compensativo.

Semplifichiamo il nostro modello a sole quattro variabili, la larghezza del marciapiede, la qualità dello spazio, la velocità dei veicoli e l’illuminazione. Un’alternativa *A* senza marciapiede (1), con ottima qualità dello spazio (5), con una velocità dei veicoli oltre i 70Km/h (1) e con un’ottima illuminazione (5), a parità di peso degli attributi, con un semplice modello basato sulla media apparterebbe alla classe “3 - Buona”. È facile intuire che questa non sarebbe certo una classificazione che verrebbe assegnata per valutare la camminabilità del segmento stradale *A*, che sarebbe piuttosto considerato dalla maggior parte degli utilizzatori – con buona probabilità – come insufficiente. È da notare che l’alternativa *A* sarebbe classificata allo stesso modo di un’alternativa *B* costituita da: un marciapiede comodo (3), una qualità dello spazio media (3), una velocità dei veicoli di 30km/h (3) e una buona illuminazione (3), che sarebbe stata quindi – ragionevolmente – classificata come buona dalla maggior parte dei pedoni.

Un metodo non-compensativo permette invece di tenere conto di valori estremi ed assegna l’alternativa ad una classe attraverso regole di assegnazione, di maggioranza o di appartenenza. Ad esempio, per l’alternativa *A* si potrebbe definire una regola che assegna l’alternativa ad una classe quando almeno il 60% dei suoi attributi appartiene a quella classe, in questo caso l’alternativa sarebbe classificata come “1 - Insufficiente”, poiché solo il 50% degli attributi è maggiore di 1. Inoltre ed oppure – a seconda della procedura adottata – si potrebbe definire una regola (di veto) per la quale le alternative senza marciapiede non possono

essere classificate oltre la prima classe, quella insufficiente. Alla luce di queste osservazioni sembra utile testare alcuni metodi non-compensativi, per capire se a partire dai dati in possesso si possa costruire un modello più accurato di descrizione della qualità dei segmenti stradali.

4.1.3.1 Dominance-based rough set

Il primo tra i metodi che possiamo testare è un metodo basato sulla teoria degli insiemi approssimati (Pawlak, 1982, 1991), si fonda sull'assunzione che elementi caratterizzati dalla stessa descrizione sono indiscernibili e formano un insieme elementare. Ogni insieme elementare descrive una parte di ciò che conosciamo dell'universo, nel nostro caso gli insiemi elementari costituiscono una regola di appartenenza del tipo *if...then*. L'insieme delle regole di appartenenza costituisce la base di classificazione delle alternative.

Quando – come nel nostro caso – si analizza l'appartenenza delle alternative descritte da variabili ordinali ad una lista di classi ordinali, diventa utile adottare un approccio che non si limiti a considerare l'indiscernibilità (l'equivalenza) tra i valori degli attributi ma che piuttosto sia basato sulla dominanza tra di essi.

In questi casi è utile applicare il così detto *Dominance-based rough set approach*⁷⁷ (da qui *rough set*), che considera nell'analisi la relazione d'ordine tra attributi e classi, restituendo delle regole di appartenenza “almeno o al più” ad una classe.

A questo punto, per non appesantire il testo, si rimanda il lettore ai dettagli metodologici in appendice B, proseguendo la descrizione dell'esperimento ipotizzando che il lettore conosca il funzionamento del metodo.

Per la definizione delle regole di appartenenza sono state usate tutte le unità statistiche, selezionando i nove indicatori risultati significativi dalle analisi di inferenza (modello B). Il valore di forza relativa minima (*Minimal rule relative strength*) per le regole è stato fissato a 10% così da ottenere un modello più esteso

⁷⁷ La procedura può essere applicata sfruttando il programma 4eMka2, sviluppato dall'*Institute of Computing Science, Poznań University of Technology* (Greco *et al.* 1997, 1999, 2001).

e meno incerto⁷⁸. Il risultato finale è di 411 regole, 258 regole certe e 153 regole approssimate; nel modello sono incluse sia le prime (quelle per cui un'alternativa è associata ad una sola categoria) sia le seconde (quelle per cui un'alternativa può essere associata a più di una categoria)⁷⁹. Le regole incerte definiscono un set di classi alle quali l'alternativa può appartenere, lasciando al decisore la scelta finale. In alcuni casi questo potrebbe essere un risultato soddisfacente, ma a noi interessa poter confrontare la qualità delle approssimazioni tra i diversi modelli adottati e per questo abbiamo necessità di un valore univoco di appartenenza per ogni alternativa. La decisione è quindi quella di associare alle alternative la classe mediana tra quelle proposte dal modello. Ad esempio, se l'alternativa *A* è classificabile come “Sufficiente, Buona o Molto Buona”, nel nostro caso sarà considerata come Buona. Quando il set di classi è pari possiamo adottare – secondo una procedura “pessimistica”⁸⁰ – il primo valore inferiore alla mediana; tra “Sufficiente e Buona” l'alternativa sarà classificata come Sufficiente, o viceversa secondo una regola “ottimistica”. La differenza tra i valori rilevati ed i valori predetti ci permette di ottenere gli scarti confrontabili con i risultati del test di regressione. La distribuzione dei residui per un modello pessimistico (Figura 4.8, sinistra) indica che il 52% dei segmenti stradali è classificato in maniera coincidente, l'89% al più in una classe di errore; per la procedura ottimistica, i risultati sono rispettivamente del 51% per le classificazioni coincidenti e del 96% per quelle al più in una classe di errore (Figura 4.8, destra).

⁷⁸ L'idea di ridurre la robustezza e diminuire l'incertezza potrebbe portare in inganno chi legge, una precisazione è necessaria. Aumentare la robustezza significa diminuire il numero di regole, scartando quelle con un rapporto tra casi simili e classificazioni differenti maggiore. Significa che maggiore è il rapporto maggiore è il numero di alternative simili che soddisfano quella regola in maniera esatta. Diminuendo la robustezza diminuisce l'efficacia delle regole ma allo stesso tempo aumenta il numero di regole, il che permette di includere più aspetti nel modello, rischiano però di rendere il modello sovra-adattato.

⁷⁹ Includere solo le regole certe aumenta il numero delle classificazioni certamente errate. È quindi preferibile una classificazione potenzialmente errata che una certamente errata. Inoltre, con valori bassi di *rule strength* le regole incerte classificano l'alternativa all'interno di un “*range certo*” nel quale una delle classi è quella corretta.

⁸⁰ Un modello può essere definito da una regola pessimista o ottimista, è pessimista quando adotta un principio di precauzione, scegliendo l'alternativa peggiore (cioè inferiore nella relazione di preferenza); mentre è ottimista quando si sceglie l'alternativa migliore (superiore nella relazione di preferenza). In questo, in base ai risultati, caso è preferibile un modello ottimistico.

Il confronto tra gli istogrammi della procedura ottimistica (Figura 4.8, sinistra) e di quella pessimistica (Figura 4.8, destra) mostra come, nel primo caso, le frequenze dei valori sottostimati siano maggiori (29% contro 20% sovrastimati), a differenza della regola pessimistica che sovrastima il 40% delle alternative e sottostima solo l'8%.

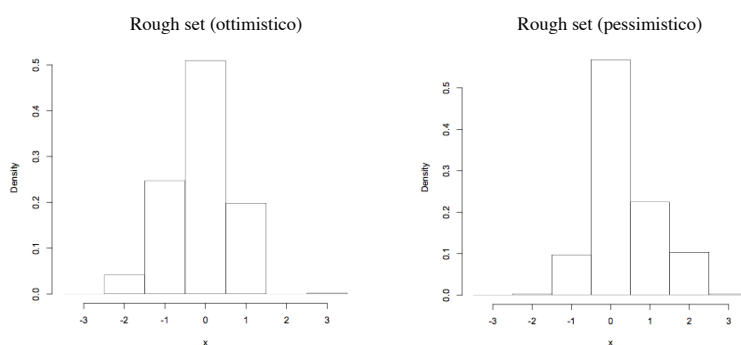


Fig. 4.8 Istogrammi dei residui della procedura *rough set*. A destra i risultati della regola pessimistica, a sinistra i risultati della regola ottimistica.

4.1.3.2 Ordinamento per regola di maggioranza

Un ulteriore approccio che vanta alcune proprietà desiderabili per il nostro modello di valutazione è la *Majority Rule Sorting procedure* (MR-Sort) (Bouyssou e Marchant, 2007a, 2007b), direttamente ispirata al lavoro di Bouyssou e Marchant è una versione semplificata dell'ELECTRE TRI (Yu 1992, Roy e Bouyssou 1993). Il principio generale dell'Mr-Sort è di assegnare le alternative ad una classe confrontando le loro prestazioni ai profili che delimitano le categorie. Un'alternativa è assegnata ad una classe superiore se e solo se è almeno buona come il profilo per una (ponderata) maggioranza di criteri, secondo una logica "di concordanza". L'ELECTRE TRI è invece una procedura di decisione, in generale, basata su due analisi complementari dette "di concordanza" e "di non discordanza". La prima misura il grado di preferenza di un'alternativa rispetto ad un'altra, la seconda dà una misura dell'inferiorità dell'alternativa migliore rispetto alla peggiore. Anche in questo caso, per approfondimenti metodologici sui due modelli, si rimanda il lettore alla descrizione dettagliata nell'appendice C.

L'interesse per il modello Mr-Sort nasce dal fatto che questo prevede alcune procedure di stima dei parametri attraverso esempi di assegnazione (Sobrie *et al.* 2013, Leroy *et al.* 2011), che – nel nostro caso – sono i giudizi e i dati dell'*audit* stradale effettuato. Nello specifico, la procedura consiste nell'individuazione del valore della soglia di maggioranza, dei pesi di ogni attributo e dei profili tra categorie attraverso la minimizzazione degli scarti tra i dati degli esempi. Un limite di applicazione di questo modello al caso studio è che la procedura è pensata per regole che il decisore può considerare correttamente classificate, o per meglio dire certe. Nel nostro caso, invece, il decisore ha espresso una valutazione per ogni segmento, indicando il giudizio di qualità percepita anche in situazioni di incertezza. Questo, aggiunto al numero elevato di attributi, non permette di ottenere risultati consistenti nella stima dei parametri a partire dal campione di dati a disposizione. Tuttavia, il modello presenta interessanti proprietà desiderabili (ad esempio, non è compensativo) e può essere applicato definendo le soglie a partire dall'osservazione dei dati "medi" ottenuti dal campione.

Il modello per il caso studio è costruito a partire dai nove indicatori che sono risultati significativi nel test di regressione lineare (modello B) e dai loro pesi normalizzati a somma 1, ottenuti a partire dai valori dei coefficienti di regressione. I pesi così definiti sono riportati in Tabella XIV.

Tab. XIV Pesi assegnati agli attributi per il modello Mr-Sort.

Attributo	X1	X2	X3	X4	X5	X7	X9	X11	X12
Peso	0,1543	0,1803	0,0772	0,0817	0,0512	0,0909	0,1039	0,0985	0,162

La definizione dei valori di soglia è stata invece prodotta a partire dalle osservazioni delle matrici di assegnazione di giudizio, in relazione ad ogni modalità di ogni variabile, assegnando l'appartenenza di una caratteristica dello spazio ad un giudizio di qualità. Per essere più chiari è utile un esempio: osserviamo la variabile "larghezza del marciapiede", è composta da cinque modalità "5 - Assente", "4 - Inadeguato", "3 - Minimo", "2 - Confortevole" e "1 - Ampio". Se guardiamo la distribuzione dei giudizi per ogni modalità otteniamo la seguente matrice Tabella XV (si veda anche il grafico a mattonelle 4.2, pag. 85):

Tab. XV Matrice della distribuzione dei giudizi.

	1 - Ampio	2 - Confortevole	3 - Minimo	4 - Inadeguato	5 - Assente
5 - Eccellente	30	6	5	0	3
4 - Molto Buono	13	31	23	2	1
3 - Buono	3	28	89	22	6
2 - Sufficiente	2	12	57	26	4
1 - Insufficiente	0	2	3	14	25

Possiamo osservare che l'assenza del marciapiede è valutata come insufficiente nel 64,1% dei casi, assegniamo perciò alla soglia b_{12} il valore di ≤ 5 . Secondo lo stesso principio di assegnazione in funzione del valore modale definiamo:

$b_{23} \leq 4$ e > 3 , $b_{34} \leq 3$ e > 2 , $b_{45} \leq 2$ e > 1 .

Analizzando ogni attributo secondo questa procedura possiamo ottenere la tabella delle soglie (Tabella XVI), dove i numeri senza parentesi indicano il valore della modalità più frequente per la classe di giudizio, mentre quelli tra parentesi indicano il valore definito successivamente, a seguito di un processo di ottimizzazione del modello. Ad esempio, per l'attributo "X9 Ciclabilità" risulta che il giudizio "Insufficiente" è preferito quando la pista ciclabile è "3 - in strada", possiamo dire che è invece più appropriato che una strada sia insufficiente se l'accesso alle biciclette è "4 - non permesso".

Tab. XVI. Valori di appartenenza ad una classe per le modalità delle variabili del modello.

	Insufficiente	Sufficiente	Buono	Molto Buono	Eccellente
X1	5	4	3	2	1
X2	5	(3)4	3	2	1
X3	4	(4)3	(2)3	(2)1	1
X4	3	3	3	2	1
X5	3	3	2	(2)1	1
X7	(5)3	(5)3	(4)3	2	(2)1
X9	(4)3	3	3	(3)2	1
X11	4	3	3	(2)1	(2)1
X12	4	2	2	2	2

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

Il valore del livello di consenso λ che definisce la soglia di maggioranza è stato fissato a 70%, sta a significare che, perché una alternativa sia classificata in una categoria, è necessario che quasi tre quarti degli attributi (pesati) supportino la tesi che quella alternativa appartiene alla classe⁸¹. Un valore elevato per la soglia di maggioranza riduce la possibilità di sovrastimare i giudizi, in questo caso il valore di λ è definito a seguito di alcuni test di comparazione tra i giudizi rilevati ed i giudizi stimati. Applicando il modello si possono comparare i valori rilevati e i valori predetti, analizzando così gli scarti tra i due giudizi di qualità. La distribuzione dei residui è riportata nell'istogramma in Figura 4.9. Approssimativamente il modello predice il 55% dei segmenti stradali in maniera corretta e classifica il 93% delle alternative con al più una classe di errore rispetto alle valutazioni sintetiche di giudizio campionate.

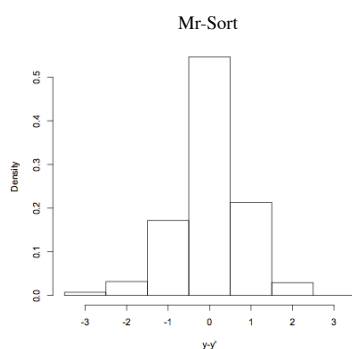


Fig. 4.9 Istogramma dei residui della procedura Mr-Sort.

4.1.3.3 Mr-Sort + *rough set* = ELECTRE TRI

Abbiamo citato il modello ELECTRE perché è il “padre” del Mr-Sort, si differenzia principalmente dal “figlio” poiché include la possibilità di valutare nel modello anche delle regole di veto. Per quel che ci interessa, possiamo intendere le regole di veto come una serie di regole decisionali del tipo *if...then* che influenzano la classificazione quando questa non è concorde alle regole (per approfondimenti sulle regole di veto si veda l'appendice C). Le regole di

⁸¹ Poiché le variabili sono discrete e non continue, il modello applicato è quello – così detto – ad “approssimazione negativa” ovvero che a parità di valore assegna l'alternativa nella classe più bassa.

classificazione ottenute nella procedura basata sul *rough set* possono perciò essere integrate nel modello Mr-Sort. Nello specifico, applicando le 411 regole si ottiene la modifica di 23 alternative, questo permette di incrementare il numero di variabili correttamente classificate a 60% ed al 94% quelle al più in una classe di differenza (Figura 4.10).

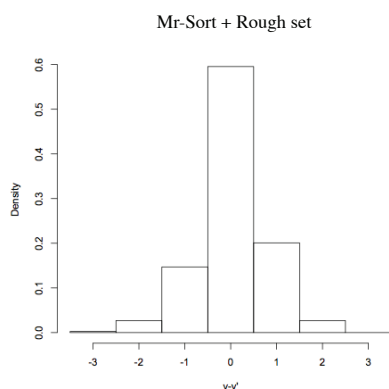


Fig. 4.10. Istogramma dei residui della procedura ELECTRE TRI.

4.1.3.4 Reti neurali artificiali

Un ultimo strumento di classificazione non-compensativo (West *et al.* 1997) che permette di inferire i giudizi di qualità dello spazio è la rete neurale.

Le reti neurali artificiali (*artificial neural networks*) rappresentano uno strumento di *data mining*⁸² utile quando la descrizione del fenomeno che si vuole modellizzare risulta essere molto complessa. La rete neurale predittiva nasce dall'intuizione di costruire dei sistemi computazionali che potessero memorizzare e modificare le informazioni, costruendo percorsi di riconoscimento e classificazione grazie all'apprendimento di informazioni, imitando il sistema neurologico del cervello umano. Secondo la definizione di Haykin (1998) una rete neurale è un processo distribuito e parallelo che ha una naturale propensione a memorizzare la conoscenza sperimentale e a renderla disponibile per l'uso. Assomiglia al cervello sotto due aspetti:

⁸² Il *data mining* è un processo di esplorazione e modellizzazione di grandi quantità di dati (*big data*) al fine di definire – attraverso procedure statistiche e algoritmiche – fenomeni, regolarità o relazioni non note. Il *data mining* predittivo permette di applicare le regole derivate dall'osservazione di un *set* dati di *training* su nuovi *set* di dati non noti.

- La conoscenza viene acquisita dalla rete tramite un processo di apprendimento.
- Le forze di connessione inter-neurale, conosciute come pesi sinattici, vengono utilizzate per memorizzare la conoscenza.

Seppur basate su principi statistici, ciò che differenzia le reti neurali dai tradizionali metodi d'inferenza è che queste non sono costituite da una rigida struttura e da un rigido insieme di ipotesi imposte al modello prima di apprendere i dati, ma la struttura delle relazioni si definisce durante l'apprendimento. Per fare un esempio, nella regressione lineare la relazione tra le variabili è imposta come lineare a priori, ciò che si testa è la “somiglianza” dei dati alla retta. Mentre nella rete neurale la funzione non è predefinita, ciò che si fa è costruire una relazione a partire da un'approssimazione dei dati.

Nonostante questi vantaggi, che aiutano nella costruzione di un modello predittivo, le reti neurali hanno lo svantaggio di una complicata interpretazione dei coefficienti, i pesi sinattici; questa procedura diventa quindi di difficile interpretazione quando si tenta di spiegare la relazione tra la variabile dipendente e le variabili indipendenti, per i quali è più appropriato l'uso di modelli statistici tradizionali. Più dettagli sul suo funzionamento sono descritti in appendice D.

Anche in questo caso sono definiti come indicatori i 9 caratteri significativi del modello di regressione lineare, così da confrontare i risultati tra i diversi modelli e perché – come abbiamo appena visto – la rete neurale si limita ad essere un modello di previsione e non di interpretazione delle relazioni.

Applicare la procedura delle reti neurali ai dati produce una percentuale di previsioni corrette del 70% ed una classificazione al più in una classe di errore del 97%. Ma a causa dell'assenza di una struttura di ipotesi rigida sulle relazioni tra le variabili, la rete neurale è una procedura che rischia di essere sovra-adattata (*overfitted*) ai dati; perciò, per applicare questa procedura, saranno necessarie alcune precauzioni che è opportuno sottolineare. Parleremo più approfonditamente del sovra-adattamento nel paragrafo 4.3.1, con test su tutti i modelli, per il momento è sufficiente una breve introduzione che è però assolutamente necessaria per confrontare i dati della rete neurale con gli altri modelli. Analizzare una rete neurale senza testare l'*overfitting* sarebbe un errore critico a causa della tenenza del modello di adattarsi ai dati, che è molto maggiore rispetto agli altri modelli per i quali – in una fase sperimentale – il test può essere eluso.

L'*overfitting* si verifica quando il modello viene costruito in maniera troppo precisa rispetto ai dati di addestramento che, come noto, contengono del rumore statistico (errore stocastico) e, nel nostro caso, incertezza nelle indicazioni del decisore che ha espresso la valutazione sintetica di qualità dello spazio. Il modello così ottenuto sarà un (quasi) perfetto modello descrittivo dei dati ma non un buon modello predittivo di nuovi dati. La struttura della rete neurale la rende soggetta a questo fenomeno, per questo è prassi comune dividere il campione in due sottogruppi, il campione di addestramento e quello di test. Perché la rete abbia un funzionamento ottimale la differenza tra le percentuali di predizione dell'addestramento e del test deve essere minima.

Per tentare di limitare al minimo la perdita di informazioni causata dalla riduzione del campione, questo è stato diviso in circa 80% addestramento e 20% test⁸³, lasciando in questa fase esplorativa l'estrazione casuale dei due insiemi al programma di calcolo. Come da prassi comune, i parametri del modello sono stati lasciati invariati, lasciandone la definizione al *software* che struttura il modello nella maniera ottimale⁸⁴. Poiché la definizione dei set è casuale ad ogni esecuzione della rete i risultati possono differire, sono stati perciò effettuati più test prendendo come riferimento i valori medi, effettuando il così detto *holdout* multiplo⁸⁵. Con questa impostazione la percentuale di correttezza in diversi test (n°10) si attesta su un valore medio di circa 60% per il campione di addestramento e 58% per quello di test, con una media degli scarti in valore assoluto ($\sum|x|/n$) del 4,7% ($\Delta_{max}=7,6\%$, $\Delta_{min}=-5,3\%$) che ci permette di concludere che l'effetto di *overfitting* è trascurabile.

⁸³ Il valore più comunemente usato è 7/3 ma, poiché il nostro campione è limitato rispetto agli standard del modello, un buon compromesso è 8/2. Altri test con valori inferiori (sino a 9/1) restituiscono valori simili, mentre l'uso dell'intero campione come training porta ad una percentuale di correttezza del 70%, che è però molto probabilmente influenzata da *overfitting*.

⁸⁴ In particolare: Il numero di livelli nascosti è 1, la sua funzione di attivazione è una tangente iperbolica. La funzione di attivazione del livello di *output* è una *softmax* e la funzione di errore entropia incrociata. Infine, il tipo di *training* è del tipo Batch.

⁸⁵ A causa del ridotto numero di unità statiche relative necessarie per questo tipo di analisi, intendiamo il nostro modello come esplorativo, lasciando test più puntuali una volta ampliato il campione. In quel caso sarà più opportuno applicare una *cross-validation* che permette di costruire gruppi di insegnamento e test in maniera sequenziale ottenendo risultati più sicuri.

Bisogna sottolineare che le reti neurali sono pensate per funzionare con campioni di dati ampi, il nostro test è quindi da considerarsi esplorativo, ma i risultati promettono buoni esiti nell'applicazione della procedura. Anticipando gli sviluppi futuri, possiamo dire che sarà certamente utile ampliare il campione ed applicare test di verifica dell'*overfitting* più minuziosi, come la *cross-validation* che sarà testata più avanti per un'ulteriore conferma (paragrafo 4.3.2).

4.1.4 Alcune considerazioni sul confronto tra i diversi metodi

A questo punto è utile fare una breve considerazione sui risultati dei diversi modelli, ci serve solo una breve sintesi, per avere chiari i risultati ottenuti sino ad ora; riprenderemo l'argomento in maniera più dettagliata e con il supporto di alcuni test di verifica, includendo anche i risultati del prossimo test, quello sugli attraversamenti (paragrafo 4.3). Per definire quale può essere ritenuto un buon metodo di classificazione dei segmenti stradali ci serve considerare tre elementi: il numero di classificazioni corrette, il numero di variabili del modello (meno sono e meglio è!) e un indicatore della variabilità, che ci indichi di quanto si discostano le classificazioni da quella corretta. Per quest'ultimo scopo esistono alcuni indicatori sofisticati (es. errore quadratico medio), ma (al momento) per semplicità il numero di alternative "al più in una classe di differenza" sembra il miglior indicatore, che dà conto della percentuale totale di classificazioni accettabili con un'indicazione di quali sono quelle che si discostano di una sola classe.

Abbiamo differenziato i metodi in compensativi e non-compensativi, dei primi la regressione ordinale è quella che ottiene i risultati migliori, mentre per i modelli non-compensativi le reti neurali sono quelle che hanno una percentuale di errore minore. Poiché per la loro complessità di calcolo le reti neurali sono più soggette ad *overfitting* rispetto agli altri modelli, nei risultati di sintesi in Tabella XVII sono riportati gli effetti del test senza *holdout* e del test che applica la procedura di verifica. Il primo modello – i cui risultati potrebbero essere influenzati da *overfitting* – hanno una chiara predominanza sulla regressione ordinale, che è invece quasi equivalente al secondo risultato delle reti neurali. Non c'è quindi una prevalenza univoca di un modello sull'altro (almeno con i dati disponibili e con i confronti che possiamo fare sino ad ora), ma possiamo dire di preferire le reti

neurali come metodo di predizione delle classificazioni perché, oltre ad uno 0,5% di precisione in più nella previsione, permettono di ottenere un buon risultato con la metà delle variabili analizzate rispetto al modello A della regressione ordinale, semplificando perciò il compito di *audit* per la raccolta dei dati.

Tab. XVII Sintesi dei risultati dei modelli di stima.

	Regressione lineare Mod A	Regressione lineare Mod B	Regressione lineare Mod C	Regressione stepwise	Regressione ordinale Mod A	Regressione ordinale Mod B	Regressione ordinale Mod C	Mr-Sort	Rough Set (pessimistico)	Rough Set (ottimistico)	Electre Tri	Reti neurali (overfitted)	Reti neurali
Correttamente classificati	55%	55%	52%	55%	62%	58%	54%	55%	52%	51%	60%	70%	62%
Al più in una classe di differenza	95%	95%	95%	96%	98%	96%	96%	93%	89%	96%	94%	97%	97%
Numero di variabili	18	9	5	11	18	9 (36)	6 (18)	9	9	9	9	9	9

Nota: i valori tra parentesi indicano il numero delle variabili in formato *dummy*

Riprenderemo questo ragionamento più avanti nel testo, includendo i risultati dei test sulle intersezioni stradali (paragrafo 4.3) e proponendo alcuni test di convalida dei risultati: un'analisi del livello di *overfitting* per i modelli di *data mining* (paragrafo 4.3.1), un test con la procedura *k-fold cross-validation* sulle reti neurali (paragrafo 4.3.2) e una verifica della bontà dei modelli (*benchmark*) rispetto ad un modello di previsione casuale (paragrafo 4.3.3).

4.2 Attraversamenti alle intersezioni

Abbiamo preannunciato che uno degli aspetti importanti nella valutazione della camminabilità sono le intersezioni stradali. Gli attraversamenti possono influenzare negativamente la camminabilità di uno spazio, questi sono il punto di incontro tra auto e pedoni e sono il luogo di maggior interazione che, come abbiamo visto, è a svantaggio del pedone. Gli attraversamenti sono generalmente

inclusi nelle valutazioni di camminabilità, secondo indici di connettività espressi sotto forma di numero di intersezioni per unità di superficie, e sono associati ad una maggiore capacità del pedone di raggiungere più destinazioni. Ciò che ne consegue è che indagini di *micro*-scala dei segmenti sono affiancate ad indici più generici che non tengono conto delle caratteristiche degli attraversamenti che, in alcuni casi, possono essere un ostacolo alla camminabilità piuttosto che un elemento di pregio.

Per questo è importante includere nel modello le componenti degli attraversamenti che – analogamente a quelle dei segmenti – sono percepite come influenti nella valutazione di qualità. Rilevare gli elementi ritenuti importanti negli attraversamenti ci aiuta ad elevare ad un grado più alto la costruzione di uno strumento di valutazione della camminabilità; infatti – come visto nel capitolo precedente (il 3°) – gli strumenti disponibili al momento non tengono conto dell'interferenza che l'attraversamento della carreggiata veicolare comporta nella percorrenza dello spazio da parte del pedone.

Lo svolgimento dell'indagine per gli attraversamenti è analoga a quella condotta per i segmenti stradali. A partire dalla raccolta dei dati con un esperimento sul campo, sono testate le stesse procedure di analisi dell'importanza degli indicatori e di modellizzazione del giudizio di qualità.

4.2.1 Costruzione dell'esperimento

Analogamente al test per i segmenti, è stata condotta una *contingent field survey* CFS in 181 attraversamenti di 45 intersezioni (Figura 4.11) nella città di Alghero. I dati sono stati raccolti nei giorni infrasettimanali 12 e 13 Maggio da 26 studenti (divisi in 14 gruppi) e sono composti da due diverse misure: (1) una verifica analitica (*audit*) di 24 caratteristiche delle intersezioni e (2) un giudizio sintetico soggettivo di attraversabilità, descritto come la facilità e la sicurezza di percorrenza. Ogni attraversamento è stato valutato e sottoposto a revisione (*audit*) due volte, per avere il dato del pomeriggio (Giovedì 12) e della mattina (Venerdì 13). In totale il campione di 360 unità statistiche è stato ridotto a 180, uno per

ogni attraversamento, mediando⁸⁶ i record che si riferiscono allo stesso attraversamento. Ugualmente alla ricerca per i segmenti stradali, ad ogni coppia di rilevatori è stato assegnato un gruppo differente di incroci per l'attribuzione della valutazione sintetica e per la raccolta degli attributi dello spazio.



Fig. 4.11 Intersezioni valutate nella città di Alghero (sinistra) ed esempi di attraversamenti (intersezioni 13 e 14) (destra).

La lista dei 24 attributi è riportata in Tabella XVIII. In questo caso si analizzano variabili quantitative sia continue che ordinali, appartenenti a tre tipi di classi di variabili: (1) caratteristiche fisiche dell'intersezione – es. geometria dello spazio – come la larghezza della carreggiata (variabile cardinale) o le condizioni delle strisce pedonali (variabile ordinale); (2) ostacoli alla visibilità, distinti per posizione (in strada, sul bordo del marciapiede, a 1 metro dal bordo, a 2 metri dal bordo), altezza (fianco, spalle, testa o più), trasparenza (trasparente, filtrato, opaco) e permanenza (temporaneo, rimovibile, non rimovibile); e (3) pratiche d'uso, come il conteggio dei flussi, la velocità e il tempo di attesa per l'attraversamento, misurate come variabili continue.

Per la valutazione sintetica soggettiva della qualità dell'attraversabilità, ai revisori è stato chiesto di esprimere il proprio giudizio valutativo e complessivo sugli attraversamenti secondo le seguenti istruzioni:

⁸⁶ Media per le variabili continue, mediana per le variabili discrete.

Tab. XVIII Elenco degli attributi (variabili) di ogni attraversamento.

Attributi (variabili)	Livelli (modalità)
X1 Larghezza carreggiata in metri	(continuo)
X2 Larghezza marciapiede (sinistra & destra) in metri	(continuo)
X3 Numero di corsie carrabili	(continuo)
X4 Senso unico di marcia	1 Sì, 0 No
X5 Corsia ciclabile	1 Sì, 0 No
X6 Semaforo	1 Sì, 0 No
X7 Scivoli per disabili	1 Sì, 0 No
X8 Attraversamento rialzato	1 Sì, 0 No
X9 Isola di attraversamento	1 Sì, 0 No
X10 Barriere verticali di indirizzamento del pedone (sinistra & destra)	1 Sì, 0 No
X11 Estensione del marciapiede	2 Sì, da entrambi i lati, 1 Sì, da un lato, 0 No
X12 Condizioni strisce pedonali	2 Yes, nitida, 1 Yes, sbiadita, 0 No
X13 Spazio di sosta e spazio per camminare (sinistra & destra)	3 Ampio (4-6 o più persone in attesa più passaggio), 2 Medio (2-3 persone in attesa più poco passaggio), 1 Limitato (1 persona in attesa no passaggio)
X14 Posizione dell'attraversamento rispetto al bordo dell'intersezione	2 In linea (meno di 10 passi dal bordo), 1 Laterale (più di 10 passi dal bordo), 0 Assente
X20 Frequenza biciclette	(continuo) [frequenza in 10 min]
X21 Frequenza automobili	(continuo) [frequenza in 10 min]
X22 Frequenza bus e camion	(continuo) [frequenza in 10 min]
X23 Frequenza moto e motorini	(continuo) [frequenza in 10 min]
X24 Frequenza pedoni	(continuo) [frequenza in 10 min]
X25 Velocità dei veicoli (effettiva) media di 5 veicoli rilevati	(continuo) [valore medio in km/h]
X26 Tempo di attesa per l'attraversamento	(continuo) [valore medio in secondi]
X30 Presenza di ostacoli alla visibilità sul bordo	1 Sì (presenza di ostacoli in strada e/o sul bordo), No (assenza di ostacoli)
X31 Presenza di ostacoli alla visibilità all'"interno"	1 Sì (presenza di ostacoli oltre a 1 metri dal bordo), No (assenza di ostacoli)
X32 "Somma dei livelli" di ostacoli alla visibilità	(frequenza) Presenza + Altezza + Trasparenza + Permanenza

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

“Per ogni attraversamento va espresso un giudizio sintetico di percezione della qualità e attraversabilità pedonale. Il giudizio è espresso su una scala *qualitativa ordinale*, da 1 (insufficiente) a 5 (eccellente), tenendo conto delle caratteristiche fisiche dell’attraversamento, dell’incrocio e della via carrabile; ed in generale di quanto l’attraversamento è *sicuro e confortevole*.

Per valutare la camminabilità *non* occorre considerare la distanza dal centro e gli elementi temporanei di disturbo (es. lavori in corso).”

Il significato dei livelli della scala (*Likert*) da 1 (insufficiente) a 5 (eccellente) va inteso nel seguente modo:

5. Eccellente: massimo confort del pedone e sicurezza elevata; è facile attraversare in tempi rapidi e lo spazio per il pedone in attesa è abbondante, l’attraversamento è regolato o strutturato in modo tale da non generare pericolo per chi attraversa.
4. Molto buona: incrocio confortevole da percorrere, con un buon livello di sicurezza; è abbastanza semplice attraversare e lo si fa in tempi brevi, lo spazio per il pedone in attesa è buono, l’attraversamento è regolato o strutturato in modo tale da rendere l’attraversamento abbastanza sicuro.
3. Buona: incrocio con confort e sicurezza nella media; l’attraversamento non è troppo difficoltoso, con tempi di attesa medi, lo spazio per il pedone in attesa è sufficiente, l’attraversamento è regolato o strutturato in modo tale da garantire un livello di sicurezza medio prestando la giusta attenzione.
2. Sufficiente: incrocio non molto confortevole e non troppo sicuro; l’attraversamento ha alcune caratteristiche che rendono il passaggio difficoltoso, lo spazio per il pedone in attesa è poco, con tempi di attesa medio-lunghi, l’attraversamento non è regolato o è strutturato in modo tale da offrire un livello di sicurezza medio prestando molta attenzione.
1. Insufficiente: incrocio per niente confortevole e insicuro; è difficile attraversare ed è necessario attendere molto tempo, lo spazio per l’attesa del pedone è insufficiente, l’attraversamento non è regolato o ha una struttura che rende pericoloso spostarsi da una parte all’altra.

4.2.1.1 Regressione lineare multipla

La prima analisi è stata condotta sulle frequenze dei giudizi valutativi. Il primo livello (5 - Eccellente) è stato assegnato solo da un gruppo di studenti, per solo l'1,7% del totale degli attraversamenti. Conseguentemente, il giudizio massimo è stato convertito nel secondo livello (4 - Molto Buono), ottenendo le seguenti frequenze: 4 - Molto Buono 16,3%, 3 - Buono 35,4%, 2 - Sufficiente 26,8%, 1 - Insufficiente 21,5%.

Un secondo *step* è stato quello di generare alcune nuove variabili, elaborando le originali registrate durante l'*audit*. Per prima cosa si è deciso di costruire alcuni indici per combinare misure simili per i due lati della strada, come “X2.3 larghezza minima del marciapiede”, o “X13.3 area di attesa all'attraversamento minima”⁸⁷.

Oltre queste sono utili alcune altre trasformazioni di variabili cardinali in variabili ordinali, per ridurre gli errori di misura, come “X25.2 velocità media dei veicoli”, diviso in 5 modalità {<5; 5÷25; 25÷40; 40÷60; >60} km/h⁸⁸.

Parte dell'*audit* è stata destinata alla rilevazione della presenza di ostacoli alla visibilità in quattro differenti posizioni: sulla strada, sul bordo del marciapiede, ad un metro dal bordo del marciapiede e a 2 metri dal bordo. Per ogni posizione e per entrambi i lati della strada – guardando verso la direzione di arrivo dei veicoli ed in posizione centrale rispetto all'attraversamento – si è osservata la presenza di ostacoli alla visibilità (1 Sì, 0 No). Se Sì, si è registrata l'altezza degli ostacoli (1 Fianchi, 2 Spalle, 3 Testa o più), la trasparenza (1 Trasparente, 2 Filtrato, 3 Opaco) e la permanenza (1 Temporaneo, 2 Rimovibile, 3 Irremovibile). Riguardo queste informazioni si sono elaborate le variabili X30, X31 e X32, rispettivamente “presenza di ostacoli sul bordo del marciapiede”, “presenza di ostacoli a più di un metro dal bordo del marciapiede” e “la sommatoria di: presenza di ostacoli, altezza, trasparenza e permanenza” che ricapitola le informazioni sugli ostacoli

⁸⁷ Sembra sensato analizzare congiuntamente alcune misure complementari, infatti – in un attraversamento – è inutile considerare un marciapiede ampio se, dall'altro lato, il marciapiede è assente. In questo caso ci pare quindi più opportuno analizzare la larghezza minima.

⁸⁸ Questo tipo di modifica dei dati si è resa necessaria a causa di un'elevata possibilità di errore nella stima della misura.

alla visibilità pedonale. Perciò, la variabile X30 (rispettivamente X31) è: “1 Sì” se la presenza di ostacoli è “1 Sì” per almeno una tra le possibilità “su strada” o “sul bordo” (rispettivamente “a un metro dal bordo” e “a due metri dal bordo”), “0 No” altrimenti. Mentre la variabile X32 è ottenuta sommando i valori ordinali delle caratteristiche per i quattro punti di vista e per il lato destro e sinistro del marciapiede, così come riportato in formula:

$$\sum_d \sum_l presenza + altezza + trasparenza + permanenza$$

Dove d indica i quattro punti di vista ed l i due lati del marciapiede.

Ad esempio, per l’attraversamento in Figura 4.12 (Via Cagliari, Alghero) si ottiene un valore totale di 20 per la presenza di alberi su entrambi i lati $2*(1[presenza]+3[altezza]+2[trasparenza]+1[permanenza])$ e per le auto parcheggiate sul lato destro $(1+2+2+1)^{89}$.



Fig. 4.12 Esempio di ostacoli alla visibilità in ‘Via Cagliari’, Alghero.

Dopo questo, normalizzate le variabili su scala da 0 a 1, sono state effettuate diverse regressioni lineari multiple per individuare il modello di correlazione tra il giudizio di qualità dell’attraversabilità (variabile dipendente) e le caratteristiche dell’attraversamento (variabili indipendenti).

⁸⁹ Anche questa formula di “somma” richiederebbe un’attenta definizione del metodo di aggregazione dei dati, sarà un elemento futuro di indagine.

La prima analisi di regressione è condotta per identificare quali variabili indipendenti hanno una significatività statistica. Il primo modello restituisce un R-quadro=0,72 (modello A in Tabella XIX). Utilizzando le variabili significative al livello 0,01 del modello A, l'analisi è stata eseguita eliminando passo-passo tutte le variabili non significative. Il modello così ottenuto (Tabella XIX, modello B) ha un R-quadro=0,66; la sua versione finale con solo variabili significative include: "X1 larghezza della carreggiata", "X6 presenza del semaforo", "X7 scivoli per disabili", "X9 isola di attraversamento", "X11 Estensione del marciapiede", "X12 Condizione strisce pedonali", "X14 Posizione attraversamento", "X20 frequenza biciclette", "X21 frequenza automobili" e "X24 frequenza pedoni". I risultati della regressione, in particolare il segno dei coefficienti di ogni variabile, conferma l'idea intuitiva che le variabili X1, X20 e X21 influenzino negativamente il valore di sicurezza e confort degli attraversamenti, mentre le altre variabili lo influenzino positivamente.

Una rappresentazione grafica della covarianza tra variabile dipendente e le variabili indipendenti significative è proposta in Figura 4.13, da diagrammi di contingenza a due vie per variabili categoria e *boxplot* per variabili continue.

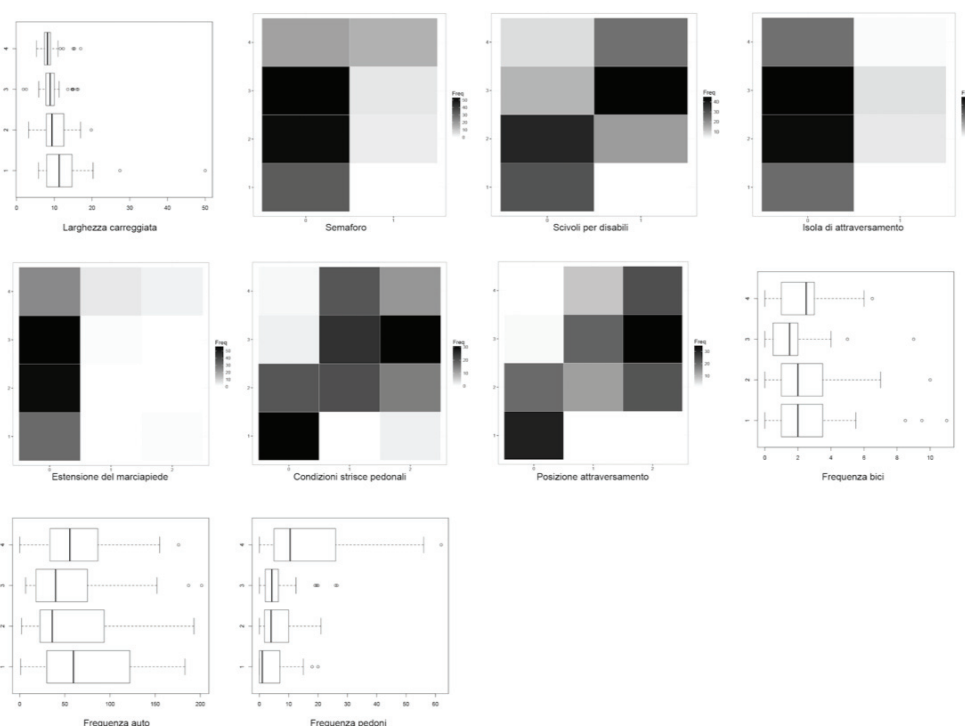


Fig. 4.13 Grafici a mattonelle e grafici a scatola (*boxplot*) della distribuzione di frequenze per giudizi e caratteristiche dello spazio.

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

Tab. XIX. Risultati della regressione lineare multipla.

	Modello A (R-quadro=0,72)				Modello B (R-quadro=0,66)				Modello C (R-quadro=0,59)			
	Stima	St. err	p-val.		Stima	St. err	p-val.		Stima	St. err	p-val.	
(Inter cetta)	0,4491	0,0763	2,5E-08	***	0,4754	0,0404	<2e-16	***	0,3304	0,0229	<2e-16	***
X1	-0,4596	0,1573	0,0040	*	-0,3533	0,1393	0,0121	*				
X2.1	-0,1312	0,0750	0,0825	.								
X2.2	-0,0653	0,1179	0,5804									
X2.3	0,2364	0,1545	0,1282									
X3	0,0983	0,1288	0,4465									
X4	-0,0167	0,0387	0,6660									
X5	-0,0089	0,0485	0,8539									
X6	0,1582	0,0469	0,0010	***	0,2083	0,0353	2,1E-08	***	0,1857	0,0379	2,2E-06	***
X7	0,0531	0,0270	0,0508	.	0,0800	0,0266	0,0030	**	0,0970	0,0279	0,0006	***
X8	NA	NA	NA									
X9	0,1076	0,0405	0,0087	**	0,1189	0,0371	0,0016	**				
X10,1	-0,0062	0,0675	0,9266									
X10,2	-0,0315	0,0655	0,6308									
X11	0,2089	0,0805	0,0104	*	0,2055	0,0698	0,0037	**				
X12	0,1580	0,0437	0,0004	***	0,1340	0,0405	0,0011	**	0,1626	0,0435	0,0002	***
X13.1	0,0969	0,0671	0,1507									
X13.2	0,0365	0,0774	0,6234									
X13.3	0,0204	0,0973	0,8344									
X14	0,1571	0,0434	0,0004	***	0,1471	0,0400	0,0003	***	0,1500	0,0426	0,0005	***
X20	-0,2098	0,0806	0,0101	*	-0,1739	0,0675	0,0108	*				
X21	-0,2014	0,0931	0,0322	*	-0,1394	0,0515	0,0075	**				
X22	0,0396	0,1091	0,7173									
X23	0,0125	0,0789	0,8739									
X24	0,3292	0,0928	0,0005	***	0,4382	0,0782	8,3E-08	***	0,4156	0,0787	3,8E-07	***
X25.1	0,2162	0,2312	0,3513									
X25.2	-0,2502	0,2351	0,2890									
X26	0,0944	0,1177	0,4239									
X30	0,0815	0,0425	0,0571	.								
X31	0,0378	0,0396	0,3413									
X32	-0,1912	0,0896	0,0345	*								

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
 Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

Anche in questo caso è utile condurre alcune regressioni singole per indagare l'importanza relativa delle variabili significative nel Modello B. I risultati delle regressioni sono riportati in Tabella XX, a partire da questi risultati possiamo notare un forte effetto singolo ($R\text{-quadro} > 0,25$) per le variabili X7, X12 e X14. Una regressione multipla per queste variabili ottiene un R-quadro di 0,46 (Modello D: intercetta 0,38 ***; X7 0,13 ***; X12 0,12 *; X14 0,20 ***). A causa della ridotta bontà di adattamento di questo modello sono state aggiunte le variabili X6 e X24 che ottengono dalla regressione semplice un $R\text{-quadro} > 0,10$. Questo ultimo modello ha un $R\text{-quadro} = 0,59$ (Modello C in Tabella XIX).

Come per il test sui segmenti, l'obiettivo di questo studio è di valutare quali caratteristiche delle intersezioni influenzano la percezione di confort e sicurezza nell'attraversamento, ma anche di sviluppare un modello che meglio permetta di stimare il valore di qualità a partire dalle informazioni sullo spazio. Per questo sono testate le differenze (i residui) tra i giudizi sintetici dei valutatori e i valori predetti dai modelli descritti. La distribuzione dei residui è mostrata in Figura 4.14 e descritta in Tabella XXI. Dai dati in figura, si può vedere che il modello A restituisce la miglior approssimazione, significa che ha i residui minori; infatti, predice circa il 64% degli attraversamenti nella classe corretta e il restante 36% in una classe di differenza. Questo significa che la probabilità è di classificare il 100% delle alternative al più in una classe di differenza. Nonostante questo, il modello B appare il preferibile, perché con 1/3 delle variabili (10 vs 30) la precisione del modello si riduce solo del 2,2% per le classificazioni corrette e del 1,7% per le assegnazioni al più in una classe di differenza. Significa che le classificazioni corrette sono il 61,7% e quelle con al più una classe di errore il 37,8%, per un totale di 99,5% di nuove classificazioni con al più una classe di differenza. La Figura 4.15 mostra una rappresentazione grafica della comparazione tra i valori registrati e i valori predetti dal modello B.

Tab. XX Regressione lineare monovariata per le variabili indipendenti più significative.

	X1	X6	X7	X9	X11	X12	X14	X20	X21	X24
Coef. est.	-0,7461	0,2748	0,2528	0,0209	0,3329	0,3381	0,3564	-0,1329	-0,1120	0,5664
Std. error	0,1929	0,0506	0,0320	0,0528	0,1061	0,0393	0,0341	0,0969	0,0743	0,1123
p-value	0,0002	1,8e-07	2,7E-13	0,6940	0,0020	3,8E-15	<2e-16	0,1720	0,1330	1,1E-06
(Intercetta)	0,7784	0,5898	0,5000	0,6234	0,6134	0,4517	0,4194	0,6538	0,6611	0,5560
(Std. error)	0,0432	0,0185	0,0226	0,0201	0,0186	0,0256	0,0246	0,0272	0,0295	0,0223
(p-value)	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16
R-quadro	0,0776	0,1420	0,2599	0,0009	0,0524	0,2941	0,3800	0,0105	0,0126	0,1251

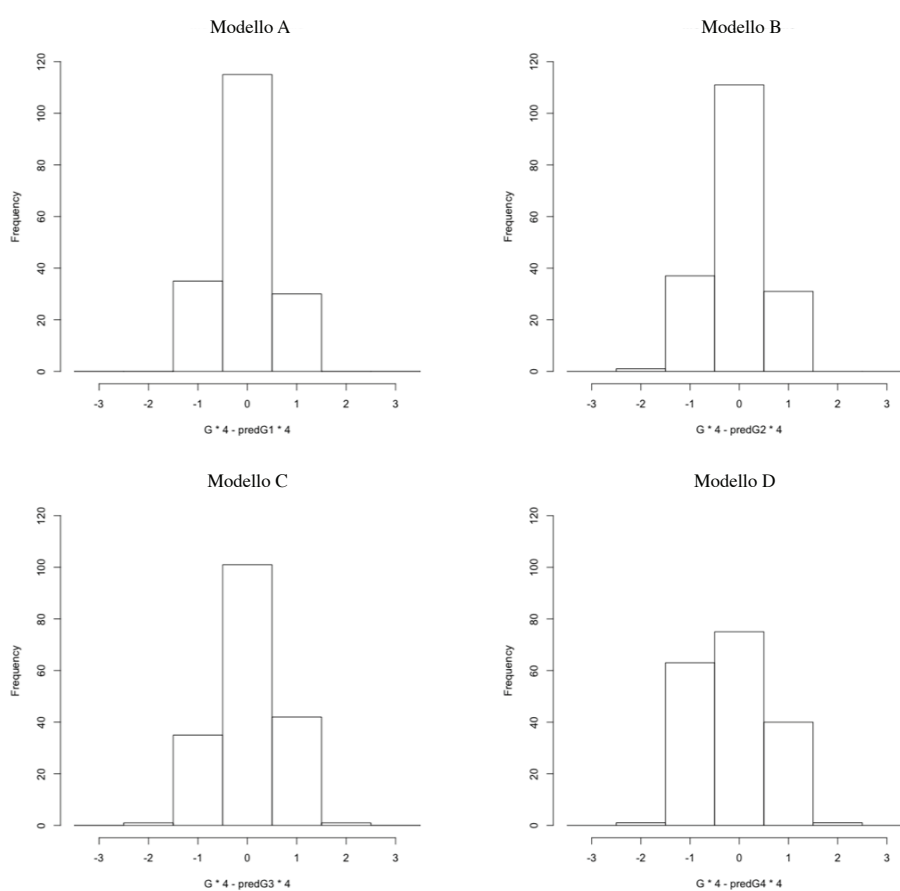


Fig. 4.14 Istogrammi dei residui regressione lineare multipla.

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
 Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

Tab. XXI Residui del modello di regressione.

	Residui					Classificazione		
	Min	1°Qu.	Mediana	3° Qu.	Max	Corretta	1 cl. off	2 cl. off
Model A	-1,50	-0,36	-0,05	0,37	1,39	63,89%	36,11%	0,00%
Model B	-1,76	-0,39	-0,05	0,37	1,38	61,67%	37,78%	0,56%
Model C	-1,54	-0,38	-0,01	0,47	1,66	56,11%	42,78%	1,11%
Model D	-1,86	-0,51	-0,11	0,49	2,49	41,67%	57,22%	1,11%

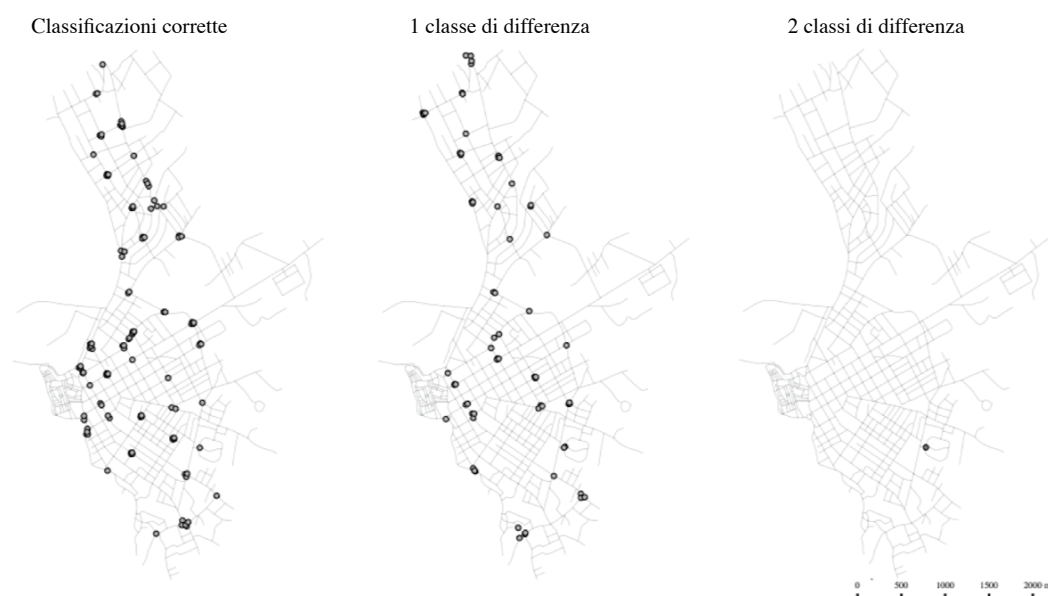


Fig. 4.15 Precisione del modello di predizione dell'attraversabilità, risultati del modello B.

4.2.2 Alcuni altri modelli di regressione

Anche per lo studio degli attraversamenti possiamo condurre diversi test con altre procedure di regressione. In questo caso ci concentreremo a testare la significatività delle variabili e i differenti modelli di stima del giudizio con la regressione *stepwise* e la regressione ordinale. Ugualmente allo studio precedente ci interessa capire se la relazione tra variabili possa essere non lineare ma, poiché molte delle variabili di questo test sono dicotomiche, testare la differenza con modelli non *probit* può essere un errore. La dipendenza tra le variabili è invece indagata con il test di correlazione; anticipando, in questo caso le variabili sono

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

molto meno dipendenti l'una dall'altra anche grazie alla costruzione (quasi ortogonale) del campione.

4.2.2.1 Regressione *stepwise*

I risultati dell'analisi *stepwise*, anche in questo caso, sono simili al modello di regressione B; con un R-quadro = 0,687 il modello include tutte le variabili del modello B più una, predicendo il 63,3% delle alternative nella classe corretta, per un totale di 99,4% di alternative con uno scarto compreso tra +1,5 e -1,5 (una classe di differenza). In Figura 4.16 è illustrato l'istogramma dei residui ottenuti dall'applicazione della retta di regressione descritta in Tabella XXII.

Tab. XXII Risultati del modello di regressione *Stepwise*.

	Stima	St. err	t value	Pr(> t)	
(Intercetta)	0,45724	0,03943	11,598	< 2e-16	***
X1	-0,36673	0,13497	-2,717	0,007275	**
X6	0,19235	0,03454	5,569	9,97E-08	***
X7	0,06006	0,02629	2,284	0,023620	*
X9	0,11797	0,03588	3,288	0,001227	**
X11	0,21591	0,06753	3,197	0,001659	**
X12	0,15691	0,03972	3,951	0,000115	***
X13.3	0,11496	0,03253	3,534	0,000528	***
X14	0,14290	0,03876	3,687	0,000306	***
X20	-0,16949	0,06518	-2,600	0,010146	*
X21	-0,14166	0,05002	-2,832	0,005191	**
X24	0,33960	0,08085	4,200	4,31E-05	***

A differenza dell'analisi dei segmenti stradali, i risultati della regressione *stepwise* introducono qualche miglioramento nelle previsioni (~1,6%) aggiungendo al modello la variabile X13.3 (area di attesa all'attraversamento minima, composta da X13.1 e X13.2). L'introduzione della variabile indipendente, che in relazione alla variabile dipendente ha un R-quadro di 0,10, non incide particolarmente sui risultati e può essere esclusa.

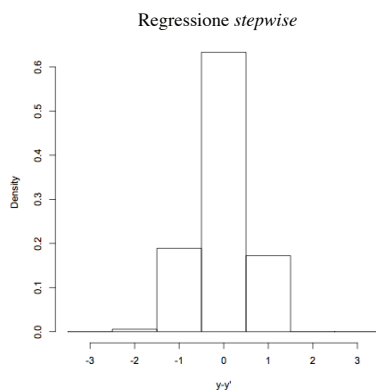


Fig. 4.16 Istogramma dei residui regressione *stepwise*.

4.2.2.2 Regressione logistica ordinale

Anche in questo caso pare utile applicare la regressione ordinale per poter analizzare la relazione di ordinalità della variabile dipendente. Anche se potessimo trattare le variabili continue come covariate è preferibile convertire in ordinali secondo i quartili, perciò in quattro classi ciascuna. Lo facciamo per ridurre le imprecisioni della misurazione (rumore) e gli errori di approssimazione intrinseci nei dati⁹⁰. Il primo modello A, con tutte le variabili, ottiene dei pseudo R-quadro di 0,770 (Cox e Snell), 0,825 (Nagelkerke) e 0,544 (McFadden) con significatività del test Chi-quadro pari a 0,000; la stima delle variabili predette è corretta per il 76,7% e al più in una classe di errore per il 99,4% delle 180 intersezioni.

Riducendo il numero di variabili nel modello B, che include quelle significative dell'analisi lineare multipla, le alternative correttamente classificate sono il 65,6%, quelle al più in una classe di differenza sono il 98,3% (Pseudo R-quadro Cox e Snell = 0,661, Nagelkerke = 0,708 e McFadden = 0,400), anche in questo caso il test del Chi-quadro appare essere significativo. Stima dei parametri e istogrammi delle due procedure sono riportati rispettivamente in Tabella XXIII e Figura 4.17.

⁹⁰ Ad esempio le velocità dei veicoli rilevate possono non essere state registrate con precisione (gli strumenti a disposizione non permettono di ottenere risultati certi), ma il loro valore approssimato fornisce comunque un'informazione sul dato reale. Un altro esempio è relativo alle frequenze, che registrate per soli 10 minuti possono non essere rappresentative della reale densità d'uso dello spazio, ma ne forniscono comunque un'indicazione approssimativa.

Tab. XXIII Risultati del modello di regressione ordinale.

	Modello A				Modello B			
	Est.	St. err	p-val.		Est.	St. err	p-val.	
[G = 1,00]	-14,476	3,72	0	***	-10,462	1,777	0	***
[G = 2,00]	-9,713	3,625	0,007	**	-8,647	1,663	0	***
[G = 3,00]	-5,742	3,544	0,105		-3,974	1,596	0,013	*
[x1=-,00]	1,126	0,861	0,191		1,061	0,567	0,061	.
[x1=1,00]	2,869	0,847	0,001	***	1,322	0,551	0,016	*
[x1=2,00]	1,995	0,735	0,007	**	0,762	0,528	0,149	
[x2.1=-,00]	1,886	1,391	0,175					
[x2.1=1,00]	0,931	0,984	0,344					
[x2.1=2,00]	1,029	0,857	0,23					
[x2.2=-,00]	0,249	1,579	0,874					
[x2.2=1,00]	0,103	1,347	0,939					
[x2.2=2,00]	1,138	1,223	0,352					
[x3=1,00]	4,205	1,924	0,029	*				
[x3=2,00]	0,921	1,511	0,542					
[x3=3,00]	0,479	1,709	0,779					
[x4=-,00]	2,815	1,008	0,005	**				
[x5=-,00]	1,41	1,11	0,204					
[x6=-,00]	-3,049	1,076	0,005	**	-2,830	0,584	0	***
[x7=-,00]	-0,042	0,516	0,934		-0,563	0,403	0,163	
[x9=-,00]	-3,422	0,87	0	***	-1,673	0,573	0,003	**
[x10.1=-,00]	0,211	1,354	0,876					
[x10.2=-,00]	0,598	1,286	0,642					
[x11=-,00]	-7,149	2,41	0,003	**	-3,122	1,447	0,031	*
[x11=1,00]	-2,444	2,934	0,405		-0,640	1,829	0,122	
[x12=-,00]	-3,814	1,99	0,055	.	-2,178	1,515	0,151	
[x12=1,00]	-1,055	0,628	0,093	.	-0,448	0,417	0,283	
[x13.1=-,00]	-2,362	1,542	0,126					
[x13.1=1,00]	-3,593	1,418	0,011	*				
[x13.2=-,00]	1,033	1,715	0,547					
[x13.2=1,00]	-0,103	1,405	0,942					
[x14=-,00]	-2,405	1,897	0,205		-2,091	1,541	0,175	
[x14=1,00]	-0,91	0,586	0,12		-0,382	0,416	0,359	
[x20=-,00]	1,9	0,783	0,015	*	1,250	0,559	0,025	*
[x20=1,00]	2,7	0,991	0,006	**	1,862	0,675	0,006	**
[x20=2,00]	1,514	0,686	0,027	*	0,956	0,487	0,050	*
[x21=-,00]	1,236	1,236	0,317		1,260	0,594	0,034	*
[x21=1,00]	0,945	1,105	0,392		1,204	0,586	0,040	*
[x21=2,00]	1,788	0,799	0,025	*	1,189	0,512	0,020	*
[x22=-,00]	1,344	0,987	0,173					
[x22=1,00]	1,966	0,945	0,037	*				
[x22=2,00]	0,76	0,808	0,347					
[x23=-,00]	-0,279	1	0,78					
[x23=1,00]	-1,48	1,063	0,164					
[x23=2,00]	0,052	0,814	0,949					
[x24=-,00]	-2,614	0,823	0,001	***	-2,214	0,530	0,00	***
[x24=1,00]	-1,575	0,757	0,037	*	-0,890	0,475	0,061	.
[x24=2,00]	-1,64	0,814	0,044	*	-1,494	0,504	0,003	**
[x25=-,00]	-0,269	0,754	0,722					
[x25=1,00]	0,478	0,706	0,498					
[x25=2,00]	0,318	0,692	0,646					
[x26=-,00]	-2,29	0,97	0,018	*				
[x26=1,00]	-2,056	0,924	0,026	*				

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
 Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

Tab. XXIII (continua)

Modello A (continua)			
	Est.	St. err	p-val.
[x26=2,00]	-0,942	0,871	0,28
[x2.3=,00]	-0,81	1,772	0,648
[x2.3=1,00]	-0,848	1,443	0,557
[x2.3=2,00]	-0,889	1,039	0,392
[x13.3=,00]	-3,476	2,071	0,093
[x13.3=1,00]	0,285	1,843	0,877
[x30=,00]	-0,143	0,688	0,835
[x30=,50]	0,516	0,577	0,372
[x31=,00]	0,51	0,613	0,406
[x31=,50]	1,921	0,675	0,004 **
[x32=,00]	-0,638	0,687	0,353
[x32=1,00]	-0,592	0,633	0,349
[x32=2,00]	-0,595	0,642	0,354

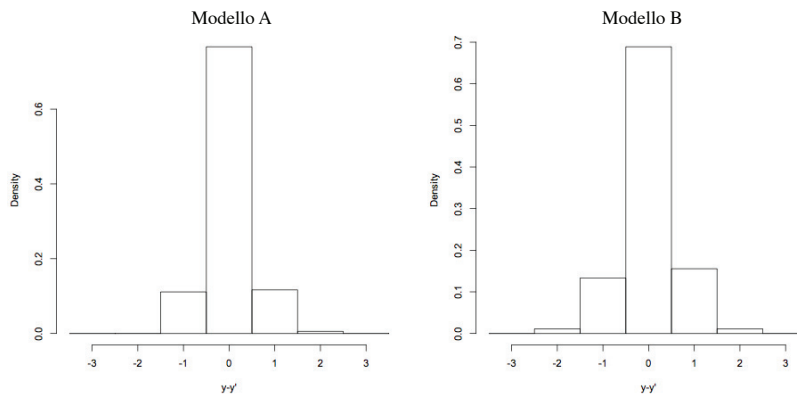


Fig. 4.17 Istogrammi dei residui regressione ordinale.

A differenza del test precedente, applicare il modello C con le sole variabili *dummy* significative – in questo caso – potrebbe essere rischioso. Infatti molte delle variabili ordinali sono già binarie, quelle che non lo sono, sono principalmente continue (ma che noi abbiamo convertito in ordinali). Convertirle in *dummy* potrebbe portare ad alcune approssimazioni eccessive che possono dare dei risultati apparentemente buoni nel confronto con il campione ma non realmente rappresentativi dell'universo.

4.2.2.3 Dipendenza tra le variabili

Anche in questo caso è utile testare l'esistenza di una relazione di correlazione tra le caratteristiche degli attraversamenti (Tabella XXIV). In questo caso si

registrano solo cinque relazioni deboli tra le variabili: “larghezza della strada” e “presenza dell’isola di attraversamento” (X1-X9), “frequenza veicoli” e “biciclette” (X20-X21) e relazione tra le tre le variabili: “presenza di attraversamenti per disabili”, “condizioni dell’attraversamento” e “posizione dell’attraversamento” (X7-X12, X7-X14, X12-X14). Sono interdipendenze che ci possiamo aspettare (es. una strada ampia ha spesso l’isola di attraversamento) e non sembrano influenzare fortemente i risultati delle regressioni. In questo caso è stato infatti possibile costruire l’esperimento scegliendo un campione (quasi) uniformemente distribuito secondo due criteri: la distribuzione spaziale uniforme tra i quartieri della città e – ciò che più conta in questo caso – una distribuzione quasi uniforme delle modalità di ogni variabile, ottenendo così dei risultati di covarianza ottimali, indicativi solo delle relazioni realmente esistenti e quindi non spuri.

Tab. XXIV. Tabella di correlazione tra le variabili significative della regressione *stepwise*.

	X1	X6	X7	X9	X11	X12	X13.3	X14	X20	X21	X24
X1	1	-0,031	-,130*	,365**	-0,07	-,148*	-0,056	-,226**	,143**	,181**	-,126*
X6		1	,229**	0,025	0,048	-0,002	,199**	,216**	,182**	,126*	0,091
X7			1	0,126	0,141	,350**	,239**	,429**	0,018	0,088	,152*
X9				1	-0,099	0,022	0,003	-0,042	0,03	,191**	-,226**
X11					1	0,104	0,1	0,055	0,065	0,106	,145*
X12						1	-0,042	,573**	-0,048	0,001	0,089
X13.3							1	0,083	0,076	0,058	,270**
X14								1	0,015	-0,027	,172**
X20									1	,408**	,168**
X21										1	-0,009
X24											1

4.2.3 Compensazione

Anche se i risultati dei test di regressione e dipendenza tra le variabili sembrano non soffrire di problemi di multicollinearità e di forte interconnessione (correlazione, dipendenza) tra le variabili, è utile testare anche in questo caso le

procedure non-compensative che ci permettono di modellizzare le preferenze secondo regole di assegnazione, regole di maggioranza e percorsi neurali.

4.2.3.1 Dominance-based rough set

Il primo test è basato sulla procedura del *rough set*, per il quale sono state analizzate le dieci variabili significative del modello a regressione (modello B). Secondo i parametri impostati nel caso precedente (*Minimal rule relative strength* = 10%) il risultato è di 719 regole, 445 certe e 274 approssimate. Seguendo la stessa procedura adottata per i segmenti stradali si ottiene una distribuzione dei residui (Figura 4.18) per cui l'87% delle alternative è correttamente classificato, il restante 13% è al più in una classe di errore.

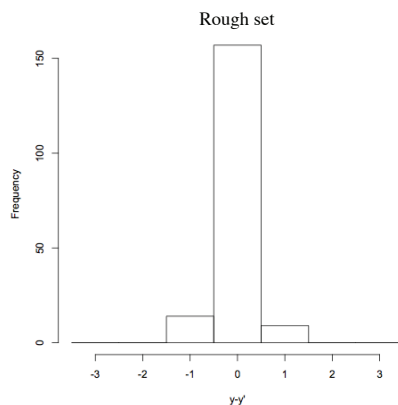


Fig. 4.18 Istogramma dei residui della procedura *rough set*.

4.2.3.2 Ordinamento per regola di maggioranza

Analogamente all'analisi dei segmenti è stata applicata la procedura non-compensativa Mr-Sort anche agli attraversamenti. Adottare la procedura in questo caso presenta un limite che, come vedremo, si riflette sul risultato del test. Infatti analizzare variabili binomiali (Sì, No) su una scala a quattro classi (Insufficiente, Sufficiente, Buono, Molto buono), con un metodo di assegnazione dell'appartenenza delle modalità ad una categoria, presenta alcuni limiti operativi. Per i dati in nostro possesso le differenze tra la prima classe "Insufficiente" e la seconda "Sufficiente" sono troppo ridotte per ottenere una classificazione precisa.

Anche in questo caso sono stati inclusi nel modello i dieci indicatori significativi della regressione lineare (Modello B) e sono stati adottati come pesi di ogni attributo i valori dei coefficienti angolari normalizzati a somma 1 (Tabella XXV).

Tab. XXV Pesi assegnati agli attributi per il modello Mr-Sort.

Attrib.	X1	X6	X7	X9	X11	X12	X14	X20	X21	X24
Peso	0,176773	0,104222	0,040028	0,059491	0,102821	0,067046	0,073601	0,087010	0,06974	0,21925

Il valore soglia per le variabili continue è ottenuto in due modi:

- per la larghezza della carreggiata sono stati scelti alcuni valori indicativi del tipo di strada, un attraversamento di 9 metri (soglia b_{34}) indica approssimativamente una strada a senso unico o a doppio senso con corsie strette (più parcheggi), uno di 11 metri (soglia b_{23}) indica una strada a doppio senso o a senso unico con due corsie strette (più parcheggi) e uno di 13 metri (soglia b_{12}) indica una strada con tre o più corsie (più parcheggi)⁹¹.
- per i valori di frequenza (auto, bici e pedoni) sono adottate tre soglie, secondo i tre valori dei quartili, primo, secondo (mediana) e terzo.

Mentre per le variabili ordinali la soglia è definita dalle frequenze di appartenenza alle classi. Ad esempio, per l'attributo "presenza di attraversamenti per disabili" la matrice di frequenze è mostrata in Tabella XXVI.

Tab. XXVI Matrice di frequenza per l'attributo "presenza di attraversamenti per disabili".

	0 - Assente	1 - Presente
1 - Insufficiente	30	3
2 - Sufficiente	38	18
3 - Buono	14	44
4 - Molto buono	8	25

⁹¹ Per corsie si intende una larghezza approssimativa di 3 metri, significa che una strada può avere due corsie da 6 metri l'una, ma verrà classificata come una strada a 4 corsie. Questo non è un errore di misura ma una proprietà ricercata; possiamo dire che la percezione di attraversamento è pressoché identica sia che si percorrano due corsie da 3 metri l'una, sia che se ne percorra una da 6. Infatti la relazione tra il giudizio ed il numero di corsie non ha prodotto risultati significativi (P -value = 0,284, $R^2 = 0,00\%$) mentre quella tra giudizio e larghezza dell'attraversamento sì (P -value = 0,000161, $R^2 = 7,71\%$).

Significa che la maggior parte degli attraversamenti con classe 1 e 2 non dispongono di attraversamenti per carrozzine, mentre la maggior parte degli attraversamenti classificati come buoni o molto buoni hanno le rampe per carrozzine. Per l'attributo X7 le alternative con rampe per carrozzine saranno perciò classificate almeno come buone, mentre quelle senza saranno indicate al più come sufficienti. La Tabella XXVII riporta le soglie definite secondo questi criteri, con alcune modifiche che appaiono essere più appropriate.

Tab. XXVII Valori di appartenenza/soglia ad una classe per le modalità delle variabili del modello.

(Freq da tabelle)	Insufficiente	Sufficiente	Buono	Molto Buono
X1	13 (3 o più corsie)	11 (2 corsie o 3 strette)	9 (1 corsia o 2 strette)	
X6	0	0	0(1)	1
X7	0	0(1)	1	1
X9	0	0	0(1)	1
X11	0	0	0(1)	1(2)
X12	0	1	2	2
X14	0	(1)2	(1)2	2
X20	3 (3° quartile)	1,75 (2° quartile)	1 (1° quartile)	
X21	91,5 (3° quartile)	47,5 (2° quartile)	24 (1° quartile)	
X24	1,5 (1° quartile)	4,5 (2° quartile)	10,125 (3° quartile)	

Prima di definire il valore del livello di consenso è utile fermarci a ragionare sulla natura delle variabili, per due motivi:

- escluse le tre misure di frequenza e la “larghezza del marciapiede” (variabili continue), le altre sei variabili sono binarie o al più a tre modalità; questo riduce la capacità di manipolazione del modello, rendendo limitate assegnazioni più puntuali.
- alcune delle variabili sono incompatibili tra loro o la presenza di una esclude la necessità di un'altra. Ad esempio, se ci troviamo su una strada ad una corsia, per questioni di geometria degli spazi, oltre che per assenza di necessità, sarà altamente improbabile trovare anche l'estensione del marciapiede e l'isola di attraversamento. Allo stesso modo se è presente il semaforo, non è (solitamente) necessario che nell'attraversamento ci sia l'estensione del marciapiede e l'isola

di attraversamento (a meno che la larghezza della carreggiata non sia molto generosa, ma non è il caso del nostro esperimento).

Date queste condizioni si rende perciò utile costruire un modello con una soglia di maggioranza relativamente ridotta, che permetta una buona classificazione alla presenza di un numero minimo di caratteristiche. È da notare che una soglia troppo alta sotto-stimerebbe eccessivamente i valori di qualità.

È stato perciò definito il valore del livello di consenso $\lambda=50\%$. La distribuzione dei residui è illustrata nell'istogramma in Figura 4.19 (sinistra). Il modello definisce con certezza circa il 47% delle variabili classificando al più in una classe di errore circa il 92% delle alternative.

Come preannunciato, una minore precisione del modello, rispetto al metodo compensativo della funzione lineare, è imputabile alla ridotta possibilità di differenziazione tra la classe “Insufficiente” e “Sufficiente”. Ipotizzando di rendere le due classi equivalenti il modello avrebbe una precisione del 61% per le alternative correttamente classificate e 93% per le alternative al più in una classe di differenza (Figura 4.19 destra).

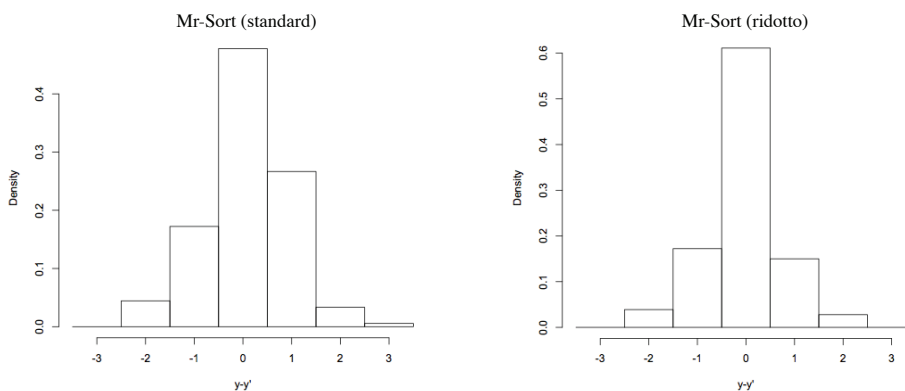


Fig. 4.19 Istogrammi dei residui della procedura Mr-Sort. Modello standard a 4 classi (sinistra) e modello ridotto a 3 classi (destra).

4.2.3.3 Mr-Sort + *rough set* = ELECTRE TRI

Seppure i risultati della procedura Mr-Sort non sembrano soddisfacenti, è comunque utile testare l'applicazione del modello ELECTRE TRI, ottenuto “sommando” alla procedura di ordinamento secondo regole di maggioranza le

regole di veto del *rough set*. In questo caso i risultati sono migliori del primo modello ma peggiori del secondo, infatti le alternative correttamente classificate sono l'83% (erano 87% nel *rough set*) per un totale di 99% di alternative al più in una classe di differenza. Del 39% di alternative che non rispettano i veti solo il 7% (~3% del totale) è costituito da incongruenza con le regole approssimate, il restante 36% è modificato da regole certe. Significa che l'applicazione dei veti certi ha avuto un forte effetto nell'incremento di qualità della misura. L'istogramma dei residui è in Figura 4.20.

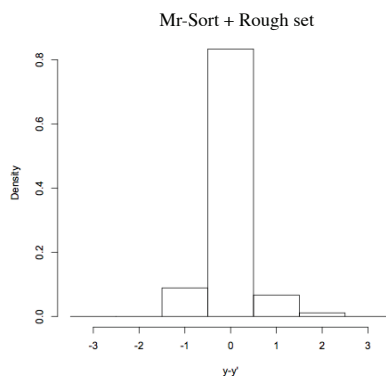


Fig. 4.20 Istogramma dei residui della procedura ELECTRE TRI.

4.2.3.4 Reti neurali artificiali

Anche per l'analisi degli attraversamenti ci sono alcuni vantaggi nell'applicare le reti neurali, abbiamo infatti visto che possiamo ragionevolmente ipotizzare l'esistenza di alcune relazioni tra le caratteristiche degli attraversamenti (es. assenza dell'isola di attraversamento quando la carreggiata ha una corsia) che la rete neurale può facilmente interpretare. La procedura di analisi dei dati è analoga a quella adottata per i segmenti così come la verifica (l'*holdout* ripetuto). L'analisi è condotta sulle 10 variabili significative ed il campione è diviso in 80% addestramento e 20% test. Anche in questo caso il modello definito dal programma ha 1 livello nascosto e le funzioni di attivazione sono la tangente iperbolica per i livelli nascosti e la *softmax* per l'output (per approfondimenti si veda l'appendice D). In media, per il gruppo di addestramento le predizioni corrette sono del 69,5% (97,1% al più in una classe di differenza), mentre quelle

del campione di test 58%, con uno scarto medio di oltre 11 punti percentuale. In test più soddisfacenti, con scarti minori, i valori di predizione corretta sono compresi tra 59,5% e 66,5%, possiamo perciò ipotizzare come ragionevole un valore intermedio tra i due. In questo caso la dimensione del campione più ristretta rispetto al caso precedente è forse insufficiente per definire una robusta rete neurale di predizione dei giudizi di qualità, che – nonostante la ridotta dimensione del campione – restituisce risultati in linea con la maggior parte dei test precedentemente condotti. Alcuni approfondimenti saranno sicuramente necessari.

4.3 Discussione sui modelli di regressione e stima dei giudizi

Abbiamo testato alcuni dei metodi che ci consentono di analizzare i dati in nostro possesso dai quali poi estrarre un modello di previsione della qualità dello spazio, ora non ci resta che definire quale è il migliore. Abbiamo già visto che il metro di misura è il numero di attributi dello spazio analizzato, la percentuale di classificazioni corrette e le classificazioni al più in una classe di differenza. Per semplicità di analisi la sintesi dei risultati è riportata in Tabella XXVIII.

In generale, rispetto ai modelli compensativi, i successi migliori si ottengono con la regressione ordinale, in particolare con i modelli B si ottengono buoni risultati con un numero abbastanza ridotto di variabili. Tra i modelli A ed i modelli B possiamo quindi sempre preferire B che (1) includono solo variabili significative e (2) ci permettono di adottare un *audit* più snello a discapito di pochi punti percentuale in meno nella precisione del modello.

Passando invece ai modelli non-compensativi i risultati sono più discordanti ma, in generale, migliori dei metodi compensativi. Possiamo quindi dire che tra i metodi compensativi e quelli non-compensativi i secondi sono preferibili oltre che per le motivazioni che abbiamo espresso in precedenza (paragrafo 4.1.3) anche per la qualità dei risultati. Nel test dei segmenti la procedura delle reti neurali sembra essere quella che restituisce i valori migliori, ma è ragionevole pensare che ci possa essere un incremento dei risultati nel *rough set* e nell'ELECTRE TRI se il campione venisse ampliato o, se più in generale, non soffrisse dei problemi di ridotta variabilità e dipendenza tra le variabili. Infatti, nel test degli

attraversamenti, escludendo i risultati sovra-adattati della prima prova per le reti neurali, il *rough set* è la procedura che pare stimare la qualità delle intersezioni con risultati nettamente superiori rispetto agli altri metodi. Vedremo però, con test più dettagliati, che anche questa procedura soffre di *overfitting* e perciò i suoi risultati non sono attendibili. Al momento, è anche in questo caso la rete neurale artificiale lo strumento che restituisce le classificazioni migliori, anche se, per giustificarlo con certezza avremo bisogno di un campione di dati maggiore.

Tab. XXVIII Sintesi dei risultati dei modelli di stima.

	Segmenti			Attraversamenti		
	Correttamente classificati	Al più in una classe di differenza	Numero di variabili	Correttamente classificati	Al più in una classe di differenza	Numero di variabili
Regressione lineare Mod A	55%	95%	18	64%	100%	29
Regressione lineare Mod B	55%	95%	9	62%	99%	10
Regressione lineare Mod C	52%	95%	5	56%	99%	6
Regressione lineare Mod D	-	-	-	41%	99%	3
Regressione stepwise	55%	96%	11	63%	99%	11
Regressione ordinale Mod A	62%	98%	18	76%	99%	29
Regressione ordinale Mod B	58%	96%	9 (36 dummy)	66%	98%	10 (34 dummy)
Regressione ordinale Mod C	54%	96%	6 (18 dummy)	-	-	-
Mr-Sort	55%	93%	9	47%	92%	10
Rough Set (pessimistico)	52%	89%	9	87%	100%	10
Rough Set (ottimistico)	51%	96%	9	87%	100%	10
Electre tri	60%	94%	9	83%	99%	10
Reti neurali (overfitted)	70%	97%	9	92%	99%	10
Reti neurali	62%	97%	9	70%	97%	10
Casuale	20%	52%	0	25%	63%	0
Casuale (distribuzioni)	25%	64%	0	27%	69%	0

Rispetto all'Mr-Sort – applicato agli attraversamenti – c'è da sottolineare che questa procedura ha una bassa precisione probabilmente a causa della binomialità della maggior parte delle variabili, non risultando quindi, di partenza, uno strumento ottimale per la stima dei giudizi. Potrebbe essere considerato un buon metodo se fosse possibile ridurre il numero delle classi di giudizio e associare gli attributi del modello in *super-indicatori* (Brüggermann e Patil, 2011, pag.98), ovvero gruppi di attributi, con l'uso dei reticoli di inclusione secondo un approccio ad ordinamento parziale e diagrammi di Hasse (Brüggermann e Patil, 2011). È un test che sarà oggetto di approfondimenti futuri.

Un'importante prova delle prestazioni dei modelli è il confronto con un classificatore casuale, per capire se gli strumenti elaborati abbiano realmente un potere di classificazione tale da renderli utili. È un test che verrà affrontato con più dettaglio in un prossimo paragrafo (4.3.3) integrando l'analisi con alcuni altri indicatori della bontà dei modelli, ad ora è utile verificare che le classificazioni corrette e quelle al più in una classe di differenza siano migliori per i modelli rispetto che per i classificatori casuali.

Come mostrato in Tabella XXVIII tutti i metodi risultano migliori sia rispetto ad un classificatore casuale con probabilità uguale per ogni giudizio, sia rispetto ad uno più raffinato che assegna le classi in base alla distribuzione osservata nel campione di dati.

Non abbiamo ancora discusso delle possibili applicazioni pratiche dei modelli (lo faremo tra poco, paragrafo 4.5), uno dei possibili sviluppi è di applicare in maniera dinamica la procedura con un continuo inserimento di dati; in questo caso possiamo pensare che sia le reti neurali sia il *rough set* possano aumentare la loro precisione incrementando – da una parte – l'accuratezza delle interconnessioni neurali e – dall'altra – la qualità delle regole con la copertura di tutti i possibili casi (combinazioni di attributi) che possono verificarsi. Inoltre, con un campione maggiore di dati si potrebbe ridurre il valore di *rule strength* del *rough set*, per scongiurare problemi di *overfitting*, includendo solo le regole più condivise.

In alcuni casi applicativi la classificazione potrebbe non per forza essere univoca, si potrebbe accettare anche di classificare le alternative entro un *range* di possibili valori di qualità. In caso di incertezza il *rough set* sarebbe un ottimo modello,

permette infatti di non classificare le alternative in una classe potenzialmente errata ma lascia la scelta finale al decisore suggerendo le classi nella quale l'alternativa può essere collocata. Questo potrebbe essere il caso in cui una classificazione errata genererebbe un problema maggiore di una classificazione incompleta o parziale (ad esempio, se dovessimo distribuire grossi finanziamenti per la riqualificazione urbana sarebbe preferibile un modello con una bassa probabilità di errore che lasci la scelta finale al decisore). Un'altra proprietà desiderabile del *rough set* è che questa procedura si presta anche ad essere applicata con più decisori (Chakhar e Saad, 2012), stimandone l'attendibilità definisce un giudizio univoco pesato utile alla procedura così come applicata al caso studio.

L'Mr-Sort – applicato ai segmenti – si è dimostrato un buon metodo di classificazione, in linea con gli altri metodi. A causa della costruzione del nostro esperimento non è stato possibile applicare la procedura di inferenza dei parametri secondo i metodi descritti da Sobrie *et al.* (2013) e Leroy *et al.* (2011) dai quali siamo partiti per la sua costruzione. Il nostro esperimento è condotto in situazioni di incertezza, dove gli intervistati hanno attribuito un giudizio di qualità dello spazio anche in situazioni di indecisione, mentre queste procedure sono strutturate per funzionare con set di assegnazioni certe. Sarebbe perciò sufficiente costruire un'indagine nella quale le alternative sono assegnate solo quando c'è una buona certezza nella loro appartenenza alla classe, così da riuscire ad ottenere risultati più consistenti. Anche questo sarà argomento di approfondimenti futuri.

Vedremo poi che un'interessante applicazione del modello è definita dall'uso dei dati *free*, come i dati di *open street map*, che hanno però il limite di essere parziali e incompleti. Una delle caratteristiche delle reti neurali è di essere un buon predittore anche quando i dati disponibili sono parziali, condizione non sufficiente per gli altri metodi testati che necessitano di un database completo. Nell'ottica di un'applicazione su scala più ampia, che non richieda un *audit* dettagliato dello spazio, le reti neurali potrebbero essere il metodo che meglio si addice a questo scopo.

Oltre alla percentuale di previsioni corrette ci sono altri parametri, che abbiamo sino ad ora ignorato, che forniscono un'accurata indicazione sulla bontà dei

modelli ed è buona prassi analizzarne. È necessario testare il livello di *overfitting* di ogni modello per capire se i valori di predizione siano sovrastimati; un metodo esplorativo è quello dell'*holdout*, che possiamo testare su tutti i modelli, uno più dettagliato è la *cross-validation*, che possiamo applicare alle reti neurali. È utile capire quale sia la bontà dei modelli nella loro capacità di includere i valori appartenenti alla giusta classe e di prevenire classificazioni errate; un modello è buono quando non ha carenze nell'assegnare le alternative ad una giusta classe e quando evita di generare classificazioni errate. È infine importante indagare se i modelli sono capaci di fornire informazioni aggiuntive rispetto ad un modello di classificazione casuale. Tutti questi test sono condotti nei tre paragrafi seguenti (4.3.x).

4.3.1 Il problema dell'*overfitting*

Nella costruzione di modelli predittivi inferiti dall'osservazione di un campione di esempi, può capitare che il modello venga “cucito” in maniera troppo dettagliata sul campione, includendo nella modellizzazione aspetti casuali (rumore), errori di campionamento o errori intrinseci nei dati. In questi casi si verifica il così detto *overfitting*, cioè un eccessivo adattamento del modello ai dati di apprendimento, che rischia di renderlo non predittivo. L'*overfitting* si verifica quando un modello comincia a “memorizzare” i dati di allenamento, piuttosto che “imparare” a generalizzare la tendenza. Questo eccesso di adattamento può derivare da differenti fattori: dal numero di parametri, dalla loro conformazione, dalla forma delle relazioni, dall'errore intrinseco, ecc. Al fine di valutare ed evitare il fenomeno dell'*overfitting* è possibile adottare alcune tecniche di castrazione del modello (es. *early stopping*, riduzione dei parametri) o convalida (es. *holdout*, *holdout* multiplo o *cross-validation* che, in linea generale, consistono nella divisione del campione in un set di insegnamento *training* ed uno di verifica *test*).

I test che abbiamo condotto sino ad ora potrebbero soffrire di un eccessivo adattamento, per questo è bene condurre un test di convalida (*holdout*) per capire quanto i modelli predicono correttamente i valori di un set di dati esterno alla loro costruzione. La pratica comune è quella di suddividere il campione di dati che si

ha a disposizione in due set, generalmente in proporzioni comprese tra il 70-30% e il 90-10%, rispettivamente per il set di apprendimento e il gruppo di test. Il primo set, definito *training set*, costituito dalla percentuale maggiore di dati è utilizzato per inferire il modello, mentre il secondo set, compreso tra il 30 e il 10% del totale è usato per verificare l'accuratezza. L'intensità dell'*overfitting* è misurata dalla differenza tra le percentuali di classificazioni corrette del set di addestramento e quello di test. Maggiore è la differenza, minore sarà la bontà del modello. Questi strumenti di apprendimento automatico svolgono molto bene il loro compito quando il campione dei dati di *training* è molto grande e quando questo non presenta rumore. Il nostro database è relativamente ristretto e, per l'applicazione del test di non accettazione *holdout*, sono state scelte come proporzioni 80% del campione per il *training* e 20% per il *test*. Il modello con i migliori risultati sarà poi analizzato più nel dettaglio con il metodo più accurato della *k-fold cross-validation*. Dei nostri modelli di apprendimento possiamo testare il modello di regressione lineare e ordinale, il *rough set* e le reti neurali. Sono chiaramente esclusi l'Mr-Sort e l'ELECTRE TRI che non sono strettamente derivati dai dati con tecniche di *data mining*. Il test è condotto estraendo il 20%⁹² dei dati dal campione in modo casuale, adottando lo stesso set *training* e *test* per ogni procedura, così da confrontare i risultati.

Per la definizione del modello di regressione con il set di *training* si è deciso di partire dai risultati dei modelli B ottenuti dai test precedenti. Per gli attraversamenti gli attributi significativi sono confermati, mentre per i segmenti le variabili "ombra e ripari" e "larghezza della carreggiata" non risultano significative. Il test è quindi condotto sia con tutte le nove variabili (Regressione lineare) sia sul modello ridotto a sette variabili (Regressione lineare 2); tutti gli altri test sono eseguiti con i dati del modello B. I risultati delle prove sono riportati in Tabella XXIX. Per entrambi gli esperimenti, le reti neurali costituiscono il modello con la differenza *delta* tra *training* e *test* minore, risultando come il metodo che soffre meno di sovra-adattamento. I dati sono confermati anche dagli studi più approfonditi condotti con *holdout* ripetuto (10 volte) (paragrafi 4.1.3.4 e 4.2.3.4) per cui la differenza tra i valori medi è di 2%

⁹² 82 unità su 408 per segmenti e 36 su 180 unità per attraversamenti.

per i segmenti e di 11% per gli attraversamenti. Se da una parte possiamo dire che per i segmenti le reti neurali sono un buon metodo di classificazione, al momento non possiamo dire lo stesso per i dati degli attraversamenti. Infatti abbiamo visto che, poiché le reti neurali necessitano di grandi quantità di dati per “imparare” le regole di classificazione, al momento i dati sugli attraversamenti non sono sufficienti per costruire un modello abbastanza robusto da dare indicazioni concrete sulla sua attendibilità. I risultati sono comunque buoni, questo induce ad approfondire più in dettaglio i risultati, un metodo è quello della *k-fold cross-validation*.

Tab. XXIX Risultati del test *holdout*. Confronto tra valori dei set di *training* e di *test*.

	classificazioni corrette			al più in una classe di differenza			
	training 80%	test 20%	Delta	training 80%	test 20%	Delta	
Segmenti	Regressione lineare	54,9%	45,1%	9,8%	95,1%	96,3%	-1,2%
	Regressione lineare 2	57,1%	46,3%	10,8%	94,8%	96,3%	-1,5%
	Regressione ordinale	58,3%	40,7%	17,6%	96,9%	95,1%	1,8%
	Rough set	54,9%	46,3%	8,6%	89,6%	90,2%	-0,6%
	Reti neurali	62,2%	58,5%	3,7%	96,9%	97,6%	-0,7%
Attraversamenti	Regressione lineare	63,9%	55,6%	8,3%	99,3%	100,0%	-0,7%
	Regressione ordinale	67,4%	61,1%	6,3%	97,2%	100,0%	-2,8%
	Rough set	83,3%	52,8%	30,5%	100,0%	91,7%	8,3%
	Reti neurali	75,0%	69,4%	5,6%	98,6%	100,0%	-1,4%

4.3.2 Rete neurale, *k-fold Cross-Validation*

Per testare con maggior dettaglio se le reti neurali possano essere un buon metodo di classificazione dei dati rilevati è buona norma eseguire un test più approfondito, la convalida incrociata. Uno dei possibili (e tra i più comuni) metodi di applicazione della procedura consiste nella divisione del set originale di dati in *k* sottogruppi uguali (*k-fold*) selezionando casualmente l'appartenenza delle

alternative ad uno dei k -set. Dei k -sottogruppi uno è adottato come *test* mentre i rimanenti $k-1$ sono usati per il *training*. La procedura è ripetuta k volte in modo che ogni k -set sia applicato almeno una volta come *test* e, di conseguenza, $k-1$ volte come *training*. Un vantaggio rispetto alla ripetizione di set definiti casualmente (*holdout* ripetuto) è che ogni osservazione è adottata sia per il *test* che per il *training*, applicandola esattamente una volta per la convalida. Come per l'*holdout*, la misura di adattamento del modello – per variabili discrete – è stimata valutando i positivi predetti correttamente⁹³; nel caso della *cross-validation*, o più in generale di misure ripetute, i valori di riferimento sono ottenuti dalla media dei risultati di ogni test. A questo punto, visti i risultati delle prove precedenti, è utile verificare con maggior dettaglio il livello di sovra-adattamento del modello a reti neurali per i segmenti stradali, scartando gli attraversamenti a causa della ridotta numerosità campionaria che, al momento, non ci permette di ottenere stime consistenti. Per un confronto con i risultati precedenti il valore di k è fissato a 5, così da ottenere *training* set composti dall'80% dei dati del set originale e *test* set composti dal 20%. Come da procedura, la definizione dei set è casuale, i risultati di ogni test sono mostrati in Tabella XXX; la media delle previsioni corrette è 60,3% (dev.st. 2,60) per il *training* e 53,4% (dev.st. 3,13) per il *test*, con uno scarto del 6,86% (dev.st. 3,25). Possiamo perciò sostenere che il modello a reti neurali è abbastanza solido per essere considerato un buon modello di predizione dei dati per il caso studio, ma un approfondimento con un campione di dati allargato può darci conferme più certe.

Tab. XXX Risultati *5-fold Cross-Validation* su segmenti stradali.

	k=1	k=2	k=3	k=4	k=5	Media
Training	58,5%	56,1%	62,2%	63,0%	61,7%	60,3%
Test	50,6%	52,8%	49,7%	56,6%	57,5%	53,4%
Delta	7,9%	3,3%	12,5%	6,4%	4,2%	6,9%

⁹³ Una misura alternativa, più comune, è il *misclassification error*, ma che non si adatta ai nostri scopi.

4.3.3 Il *benchmark*

Un ulteriore test trasversale che è utile effettuare è un *benchmark* delle prestazioni del nostro modello rispetto alle previsioni casuali. Il classificatore deve essere confrontato con un modello casuale per capire se, rispetto a questo, ci sia qualche differenza di predizione. Infatti, il modello potrebbe non essere abbastanza buono da differenziarsi da un modello ad estrazione casuale che assegna i giudizi arbitrariamente (con una probabilità $P(x)=1/n$) o uno che assegna i giudizi in funzione della distribuzione osservata nel campione di dati analizzato (con una probabilità $P(x)=fx/N$).

Prima di esaminare i risultati di un modello casuale e dei modelli testati, è necessario introdurre alcuni concetti chiave della valutazione dei classificatori attraverso l'analisi della matrice di confusione. Questa mette in relazione i valori osservati (in riga) con quelli predetti del modello (in colonna). La diagonale rappresenta i valori correttamente classificati, mentre gli altri elementi della matrice sono i valori classificati in maniera errata.

Nel nostro caso possiamo stimare la bontà del modello da tre indici: *Overall accuracy*, *Sensitivity (o producer's accuracy)* e *Precision (o user's accuracy)*.

Il primo indica il valore di previsione totale ed è il rapporto tra la somma dei veri positivi, cioè le alternative correttamente classificate (i valori della diagonale) ed il numero totale di alternative.

La sensibilità, calcolata per ogni classificazione, è il rapporto tra il numero di valori correttamente classificati come X (predizioni corrette) e il numero di valori (realmente) appartenenti alla classe X . Ovvero l'elemento XX fratto la somma degli elementi nella riga X ; può essere interpretato come “quanto la regola è capace di includere i valori positivi”.

La precisione è invece il rapporto tra il numero di valori correttamente classificati come X (predizioni corrette) e il numero di valori classificati come X (che potrebbero in realtà appartenere ad un'altra classe). Cioè l'elemento XX fratto la somma degli elementi nella colonna di X ; si può interpretare come “quanto la regola è capace di evitare classificazioni errate”, cioè di non classificare valori di X^* in X .

A dimostrazione, data la matrice di esempio 3×3 in Tabella XXXI, per la classe A il valore $Sensitivity = 5/7 = 71,4\%$, il valore $Precision = 5/8 = 62,5\%$. La precisione complessiva del modello (*overall accuracy*) è uguale a $19/27 = 70,4\%$. Ciò che possiamo estrarre dall'analisi di questa matrice è che le regole di assegnazione della classe C (*precision*) hanno una buona affidabilità (~85%) mentre quelle di B sono precise solo nel 50% dei casi. Le alternative di C sono infatti classificate correttamente (*sensitivity*) nel 91,7% dei casi, mentre quelle di B solo nel 37,5%. Le regole di assegnazione per la classe B hanno invece una bassa capacità (37,5%) di predire correttamente i reali valori di B e una capacità media (50%) di evitare classificazioni errate.

Tab. XXXI Matrice esempio della distribuzione tra giudizi previsti e osservati.

Matrice esempio		Previsti			<i>Sensitivity</i>
		A	B	C	
osservati	A	5	2	0	71,4%
	B	3	3	2	37,5%
	C	0	1	11	91,7%
<i>Precision</i>		62,5%	50,0%	84,6%	70,4%

Per includere questi concetti in un unico indice di affidabilità del modello è possibile calcolare il valore K (*Cohen's kappa coefficient*) secondo la formula seguente:

$$K = \frac{N \sum_{i=1}^k x_{ii} - \sum_{i=1}^k (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^k (x_{i+} \times x_{+i})}$$

dove N indica la somma degli elementi della matrice, $\sum x_{ii}$ indica la somma dei valori correttamente classificati (diagonale), x_{i+} indica i valori in riga della i -esima classe e x_{+i} indica i valori in colonna della i -esima classe.

In sostanza il coefficiente kappa è una misura di quanto i valori classificati dal modello sono in accordo con la classe reale, tenendo in considerazione, oltre al grado di classificazione corretta (*accuracy*), anche la *sensitivity* e la *precision* di

ogni gruppo. Il coefficiente kappa ha diverse applicazioni, è generalmente adottato in statistica per valutare la casualità delle corrispondenze tra due decisori, ma nel nostro campo – del *data mining* – può essere adottato come misura di comparazione tra algoritmi di classificazione, in ausilio ai tre indici appena descritti. Il valore è $K=1$ quando il modello è perfetto, $K=0$ quando il modello classifica i dati in maniera uniforme e $K=-1$ quando il modello classifica i dati nella classe opposta a quella di appartenenza. Per il caso in esame si ottiene:

$$K = (27(19) - ((7*8) + (6*8) + (13*12))) / (27*27 - 260) = 253/469 = 0,539.$$

Commentati questi indici possiamo analizzare la bontà dei nostri modelli rispetto ad un metodo casuale, ma partiamo da uno perfetto. Un modello perfetto, predice sempre in maniera corretta le alternative, la sua matrice di confusione, in Tabella XXXII, mette in relazione le classificazioni osservate (in riga) con le previste (in colonna). In questo caso la previsione totale (*overall accuracy*) è pari al 100% (campo in basso a destra, in grassetto), così come i valori di *sensitivity* (ultima colonna a destra) e di *precision* (penultima riga in basso), mentre kappa corrisponde a 1. In questo caso ci interessa analizzare anche la frequenza di ogni valore (percentuale globale), pari al 20% per ogni classe, descritta nell'ultima riga in basso.

Tab. XXXII Matrice della distribuzione tra giudizi previsti e osservati in un modello perfetto.

	Previsti					Percentuale correttezza (Recall, Sensitivity)	
	Perfetto	1	2	3	4		5
1	1,00	-	-	-	-	100%	
2	-	1,00	-	-	-	100%	
osservati	3	-	-	1,00	-	100%	
	4	-	-	-	1,00	100%	
	5	-	-	-	-	1,00	100%
Precision	100%	100%	100%	100%	100%	100%	
Percentuale globale	20%	20%	20%	20%	20%		

Abbiamo definito il nostro modello un buon metodo di classificazione, ma lo abbiamo fatto ipotizzando che circa il 60% di classificazioni corrette sia un risultato migliore rispetto al non avere nessuna informazione, cioè nessun modello. Questo è vero ma in realtà potremmo adottare un modello di classificazione *random*, che definisce l'appartenenza ad una delle classi secondo il lancio di un dado. La percentuale di correttezza del nostro modello deve perciò essere confrontata con quella di un modello casuale, che definisce i valori arbitrariamente secondo un'estrazione casuale delle classi di appartenenza di ogni segmento. Ipotizzando che il modello sia "non truccato", cioè che la probabilità di estrazione di ogni valore sia uguale, il nostro modello avrà una percentuale di correttezza (*accuracy*) del 20% secondo la matrice in Tabella XXXIII. Anche *sensitivity* e *precision* hanno un valore del 20%, mentre $K=0$.

Tab. XXXIII Matrice della distribuzione tra giudizi previsti e osservati in un modello casuale.

		Previsti					Percentuale correttezza
Casuale		1	2	3	4	5	
	1	1,00	1,00	1,00	1,00	1,00	20%
	2	1,00	1,00	1,00	1,00	1,00	20%
osservati	3	1,00	1,00	1,00	1,00	1,00	20%
	4	1,00	1,00	1,00	1,00	1,00	20%
	5	1,00	1,00	1,00	1,00	1,00	20%
	<i>Precision</i>	20%	20%	20%	20%	20%	20%
	Percentuale globale	20%	20%	20%	20%	20%	

Questo sarebbe un modello casuale molto elementare, potremmo affinarlo definendo la probabilità di estrazione di ogni classe in funzione della distribuzione effettiva (di ogni classe) a partire dal set di dati (Tabella XXXIV). Secondo i dati disponibili, la distribuzione dei giudizi assegnati dagli intervistati nella CFS dei segmenti stradali corrispondono a: 11% insufficiente, 17% sufficiente, 36% buono, 25% molto buono, 11% ottimo. Applicando queste percentuali alla probabilità di estrazione di ogni classe e considerando la distribuzione osservata,

la probabilità di classi correttamente classificate (veri positivi) sarà pari a $\frac{1}{4}$ (38% di falsi positivi, cioè valori sovrastimati, e 37% di falsi negativi, cioè valori sottostimati). Il valore di *accuracy* del modello aumenta a 25% a causa della distribuzione “a campana” delle frequenze dei giudizi, mentre K rimane uguale a 0. La stessa procedura per gli attraversamenti rende una precisione del modello pari a 27%.

Tab. XXXIV Matrice della distribuzione tra giudizi previsti e osservati in un modello casuale corretto.

	Casuale corretto	Previsti					Percentuale correttezza
		1	2	3	4	5	
osservati	1	5,00	7,00	16,00	11,00	5,00	11%
	2	8,00	12,00	25,00	18,00	8,00	17%
	3	16,00	25,00	53,00	37,00	16,00	36%
	4	11,00	17,00	37,00	26,00	11,00	25%
	5	5,00	8,00	16,00	11,00	5,00	11%
	<i>Precision</i>	11%	17%	36%	25%	11%	25%
	Percentuale globale	11%	17%	36%	25%	11%	

Rispetto a questi modelli casuali i nostri classificatori possono essere ritenuti migliori, garantiscono infatti una previsione molto più dettagliata di un'estrazione casuale. Ad esempio, un modello a reti neurali – che abbiamo considerato il miglior modello tra quelli testati – per il quale i set di *training* e *test* sono definiti casualmente – prevede una classificazione corretta (*overall accuracy*) del 62% per il *training* e del 63% per il *test*, secondo i risultati in matrice (Tabella XXXV).

Un limite del modello è la bassa capacità di predire correttamente i giudizi della classe 4 (*sensitivity* 32%) che sono principalmente valutati come 3 (la *precision* per la regola 3 è del 56%). Dall'analisi della matrice possiamo rilevare che il modello sovrastima il 15% delle assegnazioni e sottostima il 22% nel *training* e il 24% nel *test*; secondo un principio di precauzione è preferibile un modello che sottostima piuttosto che uno che sovrastima. Il valore di kappa per il modello a reti neurali si attesta a 0,488 per il *training* e 0,447 per il *test*, secondo alcune classificazioni

(Landis e Koch 1977, Filess 1981) questo valore può essere considerato rispettivamente come “moderato” o “discreto-buono”.

Abbiamo definito sino ad ora il modello a reti neurali il preferibile analizzando:

- l'*accuracy*, che abbiamo chiamato “classificazioni corrette”, e che indica una misura sintetica di precisione;

- le “valutazioni al più in una classe di differenza”, cioè un indicatore della dispersione.

Ma per definire con certezza il modello a reti neurali il migliore rimane da verificare che il kappa sia maggiore rispetto agli altri modelli e che la sensibilità e la precisione abbiano buoni risultati e poca variabilità.

Tab. XXXV Matrice della distribuzione tra giudizi previsti e osservati in un modello a reti neurali.

		Previsti					
	Rete Neurale	1	2	3	4	5	Percentuale correttezza
osservati (<i>training</i>)	1	23	11	1	0	1	63,9%
	2	2	42	15	0	1	70,0%
	3	1	10	88	11	5	76,5%
	4	0	6	49	28	4	32,2%
	5	0	0	3	8	26	70,3%
	<i>Precision</i>	88,5%	60,9%	56,4%	59,6%	70,3%	61,8%
Percentuale globale		7,8%	20,6%	46,6%	14,0%	11%	
osservati (<i>test</i>)	1	6	2	0	0	0	75,0%
	2	0	5	5	0	0	50,0%
	3	1	2	28	2	0	84,8%
	4	0	0	10	3	2	20,0%
	5	0	0	2	1	4	57,1%
	<i>Precision</i>	85,7%	55,6%	62,2%	50,0%	66,7%	63,0%
Percentuale globale		9,6%	12,3%	61,6%	8,2%	8,2%	

Tab. XXXVI Sintesi dei valori di *sensitivity*, *precision*, *accuracy* e di kappa.

Segmenti	<i>Sensitivity</i>					<i>Precision</i>					Kappa	<i>Accuracy</i>
	1	2	3	4	5	1	2	3	4	5		
Reg. lineare 1	51,4%	50,0%	74,6%	45,0%	24,3%	82,6%	60,0%	55,0%	42,9%	64,3%	0,377	54,9%
Reg. lineare 2 *	59,0%	35,4%	78,1%	51,0%	52,9%	95,8%	52,3%	60,1%	42,6%	94,7%	0,434	57,1%
Reg. ord. *	78,4%	16,7%	72,9%	45,0%	54,1%	49,2%	60,0%	59,7%	48,0%	60,6%	0,398	58,3%
Rough set	59,5%	77,8%	88,1%	12,5%	2,7%	100,0%	60,0%	47,3%	76,9%	100,0%	0,358	54,9%
Mr-sort †	48,6%	38,6%	84,8%	23,8%	52,8%	78,3%	55,0%	48,7%	54,1%	61,3%	0,367	53,4%
Electre tri †	64,9%	45,6%	87,5%	28,6%	55,6%	82,8%	61,9%	52,7%	64,9%	62,5%	0,438	58,9%
Rete neurale	63,9%	70,0%	76,5%	32,2%	70,3%	88,5%	60,9%	56,4%	59,6%	70,3%	0,488	62,2%
Casuale	11%	17%	36%	25%	11%	11%	17%	36%	25%	11%	0,000	25,0%

Attraversamenti	<i>Sensitivity</i>				<i>Precision</i>				Kappa	<i>Accuracy</i>
	1	2	3	4	1	2	3	4		
Reg. lineare	57,7%	67,4%	73,3%	52%	83,3%	62%	56,9%	81,3%	0,506	63,9%
Reg. ord.	69,2%	54,5%	77,1%	69,2%	75%	60%	62,7%	85,7%	0,548	67,4%
Rough set *	88,5%	90,9%	72,9%	84,6%	85,2%	71,4%	89,7%	100%	0,772	83,3%
Mr-sort †	15,4%	63,6%	47,9%	53,8%	66,7%	40,0%	51,1%	60,9%	0,269	47,9%
Electre tri † •	88,5%	81,8%	81,3%	84,6%	95,8%	80%	75%	95,7%	0,755	83,0%
Rete neurale +	78,8%	58,9%	79,3%	84,8%	68,4%	71,7%	71,9%	87,5%	0,645	75,0%
Casuale	22%	27%	35%	17%	22%	27%	35%	17%	0,000	27,0%

† derivati da regole non definite con strumenti di *data mining*, I valori in tabella sono comunque derivati dal set di *training*

* il modello soffre di un eccessivo *overfitting* ($\Delta \geq 10\%$), i risultati potrebbero non essere attendibili

• il modello è derivato dal *rough set* che soffre di un eccessivo *overfitting*.

+ il valore di Kappa da confrontare con i modelli *overfitted* è per la rete neurale 0,863

Per i segmenti, la rete neurale ottiene il punteggio maggiore di *accuracy* e kappa, non è il modello che ha il punteggio maggiore per i valori di *sensitivity* e *precision*, ma, su questi, vanta una buona linearità, se non per un deficit nella *sensitivity* della classe 4. Il *rough set* ha invece il numero maggiore di valori massimi ma un'ampia forchetta tra la qualità della predizione per le diverse classi

(*sensitivity* classe 3=12,5% e classe 4=2,7%). I suoi valori di kappa e *accuracy* sono relativamente bassi e il pericolo di *overfitting* è al limite $\Delta=8,6\%$.

Per gli attraversamenti il modello a reti neurali si configura invece il migliore da tutti i punti di vista, kappa e *accuracy* sono massimi, i valori di precisione delle regole e di capacità di evitare falsi positivi sono alti ed in linea con il valore di *accuracy*. Solo il *rough set* e il suo derivato ELECTRE TRI hanno valori abbastanza buoni ma il loro valore di *overfitting* ($\Delta>30\%$) non ci consente di analizzare i dati con certezza, i due modelli sono perciò esclusi dal confronto.

Analizzati anche questi aspetti della valutazione dei modelli di classificazione derivati da *data mining* possiamo quindi confermare che, seppure con alcuni limiti, il modello a reti neurali è quello che meglio permette di stimare i giudizi di qualità percepita della camminabilità urbana. La sintesi di tutti i modelli è proposta in Tabella XXXVI.

4.4 Trade-off tra segmenti, attraversamenti e distanze.

Abbiamo discusso, nel paragrafo 3.2, due possibili modelli di valutazione della camminabilità, che si distinguono dalla semplice misura dell'accessibilità perché integrano i tre punti fondamentali del concetto di camminabilità: le caratteristiche dello spazio (opportunità urbane), la sua percezione di qualità e le distanze. Per ora abbiamo discusso e ottenuto un metodo per modellizzare i primi due aspetti, ciò che serve capire a questo punto è qual è il *trade-off* tra la qualità dello spazio (che comprende le opportunità urbane percepite come importanti) e la distanza da percorrere. Infatti un tragitto può essere ricco di elementi attrattivi ma eccessivamente lungo, o può essere molto pericoloso ma estremamente allettante perché breve. Ciò che ci interessa è quindi capire quale sia il rapporto che intercorre tra le caratteristiche dei segmenti, le caratteristiche degli attraversamenti e le distanze dei percorsi.

Ci sono diverse procedure che è possibile affrontare, possono essere divise in due gruppi secondo due criteri: (1-1) possono basarsi sullo studio delle curve di indifferenza e delle funzioni di utilità o (1-2) su procedure statistiche multivariate applicate alla psicometria ed all'econometria come la *conjoint analysis*; possono

essere (2-1) dirette, cioè indagare la scelta tra attributi o (2-2) possono essere indirette, cioè derivare i valori di ogni singolo attributo dalla disaggregazione delle preferenze collettive.

Dei metodi che studiano le curve di indifferenza in maniera diretta troviamo ad esempio i giudizi di indifferenza (*indifference judgments*) o il metodo dell'autovalore di Saaty dell'AHP, entrambi basati sul confronto a coppie tra attributi, dai quali è possibile derivare la costruzione delle curve utilità e quindi del valore di ogni elemento della valutazione.

Ma, come per gli esperimenti condotti, ci interessa derivare in maniera indiretta la preferenza di un elemento sugli altri a partire da un giudizio complessivo che – rispetto ad una preferenza per ogni variabile – renda la valutazione più “naturale” (Furlan e Martone, 2011). In maniera indiretta, sempre riferendoci alle curve di utilità, è possibile desumere – attraverso la scomposizione delle preferenze – il valore di ogni singolo attributo ed il *trade-off* tra le componenti della misura con il metodo UTA proposto da Jacquet-Lagrèze e Siskos (1982)⁹⁴.

In breve, il metodo è basato su una regressione ordinale e una programmazione lineare. A partire da una relazione di *outranking* (*partial order*) il modello lavora su una curva di utilità additiva ($U = \sum u_i$)⁹⁵ approssimata a tratti. Le funzioni di valore marginali sono ottenute da un problema di programmazione lineare che minimizza gli errori della valutazione. I risultati permettono di stimare le curve di utilità dalle quali è possibile ricavare il valore di importanza relativo di ogni caratteristica. Un limite di questo metodo è che le funzioni devono essere monotone, cioè non discordanti, questo per il nostro tipo di analisi non sempre è facile da ottenere.

Un metodo alternativo, indiretto ma basato sulle procedure econometriche, è il *discrete choice experiment* analizzato con la *conjoint analysis*. È spesso applicato in letteratura a casi simili al nostro, come ad esempio lo studio delle preferenze di trasporto (Liao *et al.* 2015), della disponibilità a pagare per il miglioramento del paesaggio (Tempesta *et al.* 2014), per la stima del valore degli alberi

⁹⁴ Per ulteriori approfondimenti si veda anche: Jacquet-Lagrèze, E., Siskos, Y. (2001) Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research*, 130(2): 233–245.

⁹⁵ Dove U indica l'utilità totale e u_i l'utilità marginale dell' i -esimo attributo.

nell'ambiente urbano (Giergiczny e Kronenberg, 2014) o la disponibilità a pagare per tempo di viaggio risparmiato (Hensher, 2004).

Tra le diverse possibili applicazioni del metodo, nel nostro caso sarebbe possibile rappresentare graficamente le caratteristiche dello spazio per proporre una comparazione paritaria (*choice experiment*), una classificazione (*contingent rating exercise*) o un ordinamento delle alternative (*contingent ranking experiment*)⁹⁶.

In questo metodo la scomposizione delle preferenze è basata sulla *conjoint analysis* che, attraverso lo studio (statistico⁹⁷) degli effetti combinati tra caratteristiche che compongono le alternative, permette di estrapolare l'importanza relativa di ogni caratteristica e di ogni sua modalità.

Queste sono solo due delle possibili procedure che possono essere utili a definire la relazione tra gli elementi di una misura della camminabilità, questo è solo un punto di partenza.

4.5 Possibili usi dei dati e dei modelli

Abbiamo anticipato che il nostro modello può avere più usi, vediamoli! Abbiamo introdotto la nostra ricerca come uno strumento di stima della qualità dello spazio, ci sono diversi modi pratici di applicazione che possono funzionare con i nostri dati.

- Per prima cosa lo strumento può diventare un sistema di supporto alla decisione per le amministrazioni, un continuo *audit* dello spazio permetterebbe ai decisori pubblici di conoscere lo stato di qualità e definire dove intervenire per incrementare la qualità della camminabilità in riferimento agli obiettivi di equità spaziale che abbiamo discusso in precedenza.
- Valutare la camminabilità di uno spazio può essere uno strumento di verifica dell'efficacia di progetti urbani, il modello può quindi essere applicato come

⁹⁶ Per approfondimenti sulla nomenclatura e sulle procedure di analisi del *choice modeling* si veda: Hanley, N., Mourato, S., Wright, R. E. (2001) Choice Modelling Approaches: a Superior Alternative for Environmental Valuation?, *Journal of Economic Surveys*, 15(3): 435–462.

⁹⁷ Nella *conjoint Analysis* di tipo metrico, secondo il modello di preferenza *part-worth*, per la stima dei parametri si ricorre al modello di regressione multipla con variabili *dummy* basata sui minimi quadrati (*Ordinary Least Squares*). Dove le variabili di risposta *y* costituiscono la variabile dipendente e le variabili attributi *x* corrispondono alle variabili indipendenti.

strumento di previsione degli effetti indotti dalla modifica della forma e dell'uso dello spazio pedonale.

- Una mappa dell'accessibilità può anche essere uno strumento per i pedoni, un *google maps* della camminabilità che definisce il miglior percorso per un pedone non limitandosi al conteggio delle distanze come (ad oggi) avviene per i navigatori satellitari pedonali.
- La camminabilità di un ambiente urbano è certamente un valore aggiunto al valore degli immobili, è un'esternalità positiva che influenza le scelte abitative e influenza indirettamente i valori di mercato. Uno strumento di valutazione della camminabilità, in particolare uno orientato alla misura di camminabilità verso possibili destinazioni diventa uno strumento di complemento al calcolo del valore immobiliare degli edifici o di supporto alla scelta della zona di residenza.
- La camminabilità non è uguale per tutti, abbiamo parlato dei *non-cittadini* e di come lo spazio debba essere differente per offrirgli le stesse capacità dell'*uomo-tipo*. Il nostro strumento, incrociato con l'appartenenza ad una delle classi di individui, può essere utile per costruire delle mappe di accessibilità differenziate per popolazioni e per comprendere quali sono gli elementi percepiti come maggiormente importanti per chi ha difficoltà di movimento nello spazio. Approfondire questo tema ci porterebbe a possibili test e procedure di applicazione. Per prima cosa potremmo capire (come abbiamo fatto con la regressione lineare) quali sono gli attributi dello spazio che influenzano la percezione e capire in modo diretto quali sono le differenze tra i diversi individui. Applicando il modello per la costruzione di mappe si potrebbero confrontare le differenze di accessibilità spaziale nelle varie zone della città per i diversi abitanti, per confronti spaziali e successive elaborazioni (es. mappa della camminabilità minima percepita).
- Infine, con una mole di dati rilevante, avremmo la possibilità di testare la significatività dei diversi risultati di percezione tra gruppi di popolazione, per verificare se esiste una differenza statistica tra i giudizi. Potremmo capire quanto una popolazione è più svantaggiata (o meglio, percepisce lo svantaggio) rispetto all'*uomo-tipo* o alle altre popolazioni.

In molti dei campi di applicazione del processo di valutazione della camminabilità proposto è utile un continuo aggiornamento dei dati. In termini di giudizio di qualità percepita, più sono le valutazioni maggiore sarà la precisione del valore di riferimento e meglio ci sarà possibile fare confronti tra popolazioni differenti. Dal punto di vista dell'*audit* abbiamo invece, oltre ad una plurima sicurezza della misura, una serie storica dei cambiamenti ed un aggiornamento continuo delle modifiche che avvengono nello spazio. Possiamo quindi costruire uno strumento partecipativo, inclusivo, economico, permettendo agli abitanti di integrare il database di dati attraverso l'uso di un applicativo *web-based* mettendo a disposizione i dati per le diverse proposte che abbiamo appena elencato.

In generale possiamo vantare uno strumento che si inserisce in molti dei campi *mainstream*, quale la partecipazione 2.0, le città sane, le città *smart*, l'equità urbana e – ovviamente – la mobilità dolce.

Conclusioni

Qual è una buona strada per camminare? È meglio un *boulevard* o una *lane*? Ce lo siamo chiesti 147 pagine fa ed ora abbiamo qualche strumento in più per capirlo. Abbiamo dedotto che non è solo la lunghezza che conta, ma ciò che dobbiamo fare è guardare anche alle caratteristiche quantitative e qualitative dello spazio ed a percezioni, bisogni, necessità e caratteristiche di ogni individuo.

Il *boulevard* ha quindi il pregio di essere un viale alberato e ricco di servizi, ha generalmente una buona attrattività urbana e marciapiedi larghi, ma è costituito da almeno quattro corsie, anche se la separazione tra i pedoni e i veicoli è generalmente buona. Una *lane* non ha invece elementi di pregio ambientale e urbano, ha marciapiedi stretti e non ha servizi e attività, ma ha poche corsie carrabili e bassi flussi di traffico.

È chiaro che ogni *boulevard* e ogni *lane* sono diversi e non possiamo dare una risposta certa, ma se ci basiamo su questi due standard, con queste caratteristiche, secondo le preferenze dei nostri decisori, a colpo d'occhio il *boulevard* è uno spazio più piacevole da camminare perché ha marciapiedi ampi e attrattività urbana, che sono gli elementi desunti come più influenti. Ma anche le caratteristiche e le necessità contano, se quindi chiedessimo a dei non-cittadini, come i bambini, preferirebbero probabilmente una *lane*, a loro importa che lo spazio sia giocabile, non ricco di servizi o di buona qualità. È per questo motivo che l'approccio in questo testo è basato sulla teoria della capacità, per poter cogliere questi aspetti di ogni gruppo di individui così da costruire una città che sia a misura di tutti, cioè allo stesso tempo attrattiva ma anche giocabile, accessibile a chiunque e a misura di pedone. Ad ogni modo non ci dobbiamo preoccupare di trovare un ordinamento stretto (transitivo, antisimmetrico e completo) o largo (transitivo, antisimmetrico, riflessivo e completo) tra tipi di strade, questo testo ha proprio lo scopo di individuare una procedura automatica di classificazione (ordinamento largo) di ogni strada in base alle sue caratteristiche ed in base alle preferenze di diverse classi di cittadini.

Ma questo testo non dà risposte solo a questa domanda, ci sono alcuni altri contenuti che vale la pena discutere, procediamo con ordine.

Abbiamo introdotto il concetto di camminabilità come uno degli elementi della giustizia urbana e, tra i diversi motivi per cui lo spazio urbano deve essere camminabile, il diritto alla città e l'approccio delle capacità sono le due teorie che ci permettono di giustificare una promozione della camminabilità urbana.

Perciò leggiamo la camminabilità non solo come un aspetto della mobilità, ma lo facciamo piuttosto dal punto di vista della qualità della vita di cittadini e non-cittadini. La camminabilità è una caratteristica che genera capacità ed è vista come uno strumento per incrementare l'uguaglianza tra i cittadini, oltre che come un elemento necessario a produrre una città come *opera*. La giustizia urbana è quindi misurata non solo attraverso la componente dell'accessibilità, ma secondo il concetto di camminabilità che integra gli aspetti delle distanze, dalla disponibilità (e qualità) delle risorse e dalle capacità individuali di accesso allo spazio.

Per giungere a questo punto, nel testo abbiamo costruito una tassonomia, assente in letteratura, dei differenti approcci alla misura dell'accessibilità intesa in un'ottica di giustizia, analizzando le definizioni di *equità*, *equità urbana* ed *accessibilità*. Diverse interpretazioni di questi concetti portano ad una misura orientata alle caratteristiche dello spazio o alle caratteristiche della persona. Rispetto alle misure di accessibilità come indicatore della giustizia urbana, l'approccio in questo testo ha il vantaggio di intendere la valutazione contemporaneamente da una prospettiva *place-based* e da una *people-based*. In più, rispetto alle misure di camminabilità, questo studio considera lo spazio camminabile non solo in base alle caratteristiche dei segmenti stradali ma anche in funzione dell'attraversabilità alle intersezioni: (1) con una misura basata su un'elevata scala di dettaglio che analizza sia gli aspetti qualitativi sia quelli quantitativi del *design* urbano (indagati con un *audit*), (2) secondo la più dettagliata delle misure di accessibilità – la *network analysis* – ed (3) aggregando gli indicatori con una procedura non-compensativa e inferita da esempi di assegnazione, che permettono inoltre di validare e calibrare il modello di stima della camminabilità. Questo passaggio è condotto attraverso il disegno di alcuni esperimenti empirici di investigazione della relazione che intercorre tra le caratteristiche dello spazio e la sua percezione di qualità. I risultati hanno il vantaggio – rispetto agli studi della letteratura – di includere nell'analisi (1) un numero di caratteristiche dello spazio maggiore, (2) una descrizione più accurata

(in termini di modalità delle variabili e di oggettività delle modalità) e (3) un'analisi con più metodi di regressione e di previsione, sia basati sulla somma pesata (compensativi), sia su procedure di classificazione per regole di maggioranza, regole di assegnazione e percorsi non lineari (non-compensativi). La costruzione degli esperimenti ha avuto due obiettivi, il primo è stato quello di identificare quali sono gli attributi dello spazio che influenzano la percezione di qualità, la seconda è di ottenere una raccolta di dati utili da cui inferire i parametri dei diversi modelli di misura testati.

Rispetto al primo obiettivo, in accordo con la letteratura, sono adottate delle procedure di regressione per derivare il rapporto di causa-effetto del giudizio di qualità percepita rispetto alle caratteristiche fisiche (qualitative e quantitative) e alle pratiche d'uso dello spazio. Per i segmenti stradali sono risultati significativi nove attributi, mentre per gli attraversamenti dieci. I valori della bontà di adattamento delle procedure di regressione adottate, in linea con la letteratura di riferimento, ottengono dei valori di R-quadro compresi tra 0,59 e 0,69, rispettivamente per segmenti e attraversamenti, che in questo campo applicativo delle scienze sociali possono essere considerati dei buoni risultati. I valori dei coefficienti di ogni variabile ottenuti dalla regressione possono inoltre essere interpretati come un fattore d'importanza (peso) di ogni attributo, secondo i quali è possibile ordinare gli attributi per importanza in:

- Attrattività ambientale e urbana, illuminazione⁹⁸, larghezza del marciapiede, ciclabilità, separazione tra pedoni-veicoli, larghezza della carreggiata, sosta, densità di servizi e attività, livello di ombra e ripari per i segmenti;
- Frequenza dei pedoni, larghezza della carreggiata, presenza del semaforo, estensione del marciapiede, frequenza delle biciclette, posizione dell'attraversamento, frequenza dei veicoli, condizioni delle strisce pedonali, presenza dell'isola di attraversamento e presenza di strutture per disabili agli attraversamenti (coppie di scivoli).

Riguardo la seconda finalità degli esperimenti, l'individuazione dei modelli di previsione della qualità dello spazio, sono testate procedure sia compensative che non-compensative; come abbiamo anticipato, queste ultime sono inesplorate nella

⁹⁸ Questa variabile soffre di ridotta variabilità, il suo valore d'importanza potrebbe essere minore.

letteratura di riferimento. I metodi compensativi sono basati sulle procedure di regressione (lineare multipla, *stepwise* e ordinale), i risultati indicano la capacità del modello di predire correttamente le classi di appartenenza che – se consideriamo i modelli con sole variabili significative – sono pari a 55% di classi correttamente classificate per i segmenti e 63% per gli attraversamenti. Questi classificatori sono fondamentalmente basati su una somma pesata (coefficienti moltiplicati per variabili) e non permettono di tener conto di quegli aspetti di non-compensazione che abbiamo detto essere desiderabili nelle decisioni in campo urbanistico. Oltre questo aspetto ce n'è uno più importante, i risultati dei metodi compensativi sono mediamente inferiori rispetto a quelli ottenuti dai modelli non-compensativi, che si attestano ad un intervallo compreso tra 51% e 62% per i segmenti e 47% e 70% per gli attraversamenti.

Tra i metodi testati il *rough set* definisce le classi inferendo dal campione di dati delle regole di assegnazione, l'Mr-Sort (così come l'ELECTRE TRI) definisce la classificazione in base alla coalizione dei criteri secondo una regola di maggioranza e le reti neurali assegnano una classe di appartenenza alle diverse alternative in base a percorsi appresi dall'osservazione dei dati. Tra questi, le reti neurali permettono di ottenere una percentuale di classificazioni corrette maggiori, pari al 62% per i segmenti e 70% per gli attraversamenti, attestandosi anche come il modello che soffre meno di *overfitting*, verificato sia con il test dell'*holdout* (semplice e multiplo) sia con la *k-fold cross-validation*.

Le reti neurali, così come tutti i modelli testati, superano anche il test che abbiamo definito *benchmark*, cioè il test delle prestazioni dei modelli rispetto ad alcuni possibili classificatori casuali che avrebbero il vantaggio di definire l'attribuzione delle classi senza la raccolta di informazioni sullo spazio (*audit*). Rispetto a questi classificatori i nostri modelli assegnano la giusta categoria più del doppio delle volte, rendendosi effettivamente funzionali alla valutazione.

Lo strumento di misura individuato può quindi (1) essere applicato così come è nelle funzione di classificatore della qualità dello spazio, (2) può altrimenti essere integrato nel modello di misura della camminabilità attraverso l'implementazione dell'analisi di *network* o (3) può essere inglobato in altri strumenti di misura a supporto di varie misurazioni, come ad esempio la stima immobiliare (in questo senso la camminabilità è un'esternalità positiva!).

La misura della camminabilità è anche utile per il supporto alla decisione nelle scelte di mobilità urbana, nella gestione della città e nella valutazione di progetti urbani, il suo utilizzo può essere quello di disegnare mappe della camminabilità a supporto delle amministrazioni, per le scelte di gestione urbana o per identificare le differenze tra città o tra quartieri per le diverse categorie di individui. L'implementazione in altri strumenti complessi può essere varia (non poniamo limiti alla fantasia, ma) ad esempio, può essere orientata al supporto della navigazione pedonale, delle scelte localizzative, della misura della qualità urbana o, come accennato, della stima immobiliare.

Appendice A: I modelli ad equazione strutturale

Nelle scienze sociali che studiano il comportamento, soprattutto quando questo è osservato, è raro avere la possibilità di effettuare esperimenti in cui una singola variabile indipendente è manipolata per osservare la variazione della variabile dipendente; quando questo è possibile si sta spesso adottando una *proxy* o un indice della variabile di interesse teorico. Oppure, si possono verificare situazioni di causalità indiretta che se non incluse nel modello di analisi possono portare a correlazioni errate. Per ovviare a questi inconvenienti i modelli ad equazioni strutturali rappresentano una famiglia di metodi per l'analisi dei comportamenti che consentono di studiare le relazioni causali dirette, reciproche, indirette e condizionate tra variabili sia misurabili che latenti. Questi modelli consentono di affrontare tali problematiche tramite l'approccio LISREL (*Linear Structural RELationship*), sviluppato negli anni '70 dallo statistico-pricometrico Karl Jöreskog per stimare i coefficienti dell'analisi fattoriale, che associato ai lavori svolti da Sewell Wright negli anni trenta sulla *path analysis*, permette di studiare i legami causali tra variabili. Il LISREL proviene quindi da una duplice tradizione scientifica: quella psicometrica dalla quale ha ereditato il concetto di variabile latente e quella econometrica dalla quale ha assunto la rete di relazioni causali tra le variabili. In breve, i modelli ad equazioni strutturali permettono, a partire dalla matrice di covarianza o correlazione delle variabili, di ipotizzare un modello di relazione tra le variabili e validare con un test di ipotesi (massima verosimiglianza) se il modello è una buona rappresentazione delle possibili relazioni tra variabili e se i risultati delle equazioni di relazione (coefficienti di relazione) siano accettabili.

I Path models

I "modelli a percorso", studiati secondo l'approccio LISREL, sono caratterizzati da una serie di equazioni (strutturali) che indicano in forma matematica le relazioni alla base del processo di test del modello e possono essere rappresentati graficamente dal *path diagram*.

Per semplicità espositiva è utile partire dalla descrizione del diagramma che rappresenta la relazione tra le variabili, per arrivare poi a descrivere le equazioni ed il metodo di analisi. I diagrammi adottati nella *path analysis* si chiamano *path diagram*.

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

I.I *Path diagram*

Un metodo di rappresentazione della relazione tra le variabili e delle loro equazioni simultanee nella *path analysis* è il *path diagram*. La sua struttura è fondata sui criteri seguenti:

- Le variabili latenti o non osservate sono rappresentate da un cerchio o ellisse.
- Le variabili osservate sono rappresentate da un quadrato o rettangolo.
- I termini di disturbo, errori stocastici, sono rappresentati dalla lettera corrispondente e collegati alla variabile da una freccia lineare.
- Una freccia lineare ad una sola direzione rappresenta un'associazione causale tra le due variabili collegate.
- Una freccia curva a due direzioni rappresenta un'associazione non causale tra due variabili, ovvero una semplice correlazione.

Le variabili del modello sono catalogate in esogene ed endogene a seconda delle relazioni che hanno nel modello. Una variabile esogena (esterna) ha la caratteristica di essere sempre e soltanto la funzione di variabile indipendente, cioè che causa uno o più effetti sulle altre variabili senza essere influenzata da nessuna delle variabili del modello. Le variabili endogene (interne) sono quelle che possono svolgere sia la funzione di variabile dipendente che di variabile indipendente: possono perciò essere influenzate da una variabile e a loro volta influenzare una terza variabile, o semplicemente essere influenzate da una o più variabili. C'è da tener presente che “il *path diagram* che va ricercato è quello costituito dal minor numero di variabili collegate dal minor numero di frecce che possano descrivere il fenomeno, poiché rappresenta la spiegazione più parsimoniosa del fenomeno in esame.” (Loehlin 2004, p.6)⁹⁹.

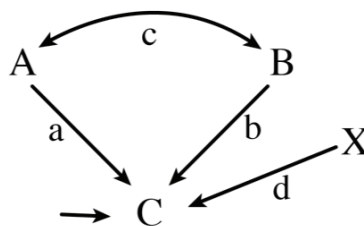


Figura A1. Esempio di *path diagram*.

⁹⁹ “The smallest number of variables connected by the smallest number of arrows that can do the job is the path diagram to be sought for, because it represents the most parsimonious explanation of the phenomenon under consideration” Loehlin (2004, p.6), traduzione propria.

Esempio 1

Nel diagramma in Figura A1 le variabili A, B ed X sono assunte come esogene ed hanno un effetto causale sulla variabile endogena C. Le variabili A e B sono assunte correlate tra loro.

I.II Path analysis

A partire dal *path diagram* che restituisce una descrizione visiva delle possibili relazioni, la *path analysis* permette di ottenere valori numerici che indicano la relativa forza di correlazione o di influenza causale a partire da dati empirici.

Il metodo sul quale si basa questa procedura è di esprimere la covarianza o la correlazione tra due variabili come somma lineare dei percorsi che legano le due variabili, tramite i coefficienti di ogni relazione del percorso, detti *path coefficients* o *structural coefficients*.

Per definire questi valori si fa riferimento alle tre regole di Wright, definite dal genetista Sewall Wright all'inizio degli anni '20:

- Non sono ammessi loop, un percorso di relazione indiretta tra due variabili non deve transitare due volte una stessa variabile intermedia.
- Non si può procedere in avanti e poi in dietro, non è consentito un percorso tra due variabili che hanno influenza su una medesima variabile se il percorso passa per la variabile dipendente.
- Non è consentito definire un percorso che attraversa più d'una freccia curva (di correlazione) tra variabili.

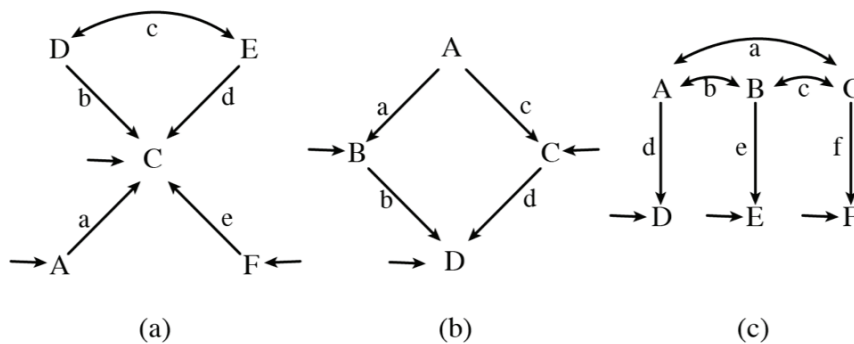
Esempio 2

Dato il diagramma in Figura A2(a) il percorso ACF rispetta le regole definite, mentre il percorso ACDECF non può essere considerato in quanto non rispetta la prima regola, la variabile C è coinvolta duplicemente nel percorso.

La seconda regola è dimostrabile nel diagramma in Figura A2(b), il percorso BAC è legittimo, mentre BDC non lo è. Nel primo caso si va indietro nella freccia BA e poi avanti in AC; il che è ammesso. Mentre nel secondo caso si procede prima da BD in avanti e poi indietro da D a C.

La terza regola è illustrata in Figura A2(c), il percorso DABE è consentito per collegare D con E, mentre DACBE, che percorre due frecce curve, no.

Perciò, seguendo le regole, per la relazione tra D ed F in Figura A2(a), i percorsi



possibili sono DCF e DECF. Il valore di r_{DF} sarà perciò la somma dei due coefficienti dei percorsi $be + cde$.

Figura A2. Illustrazione delle regole di Wright.

I.III Soluzioni numeriche del *path diagram*

La dimostrazione in ultima analisi è utile per comprendere il funzionamento operativo di una *path analysis*. Infatti, come appena visto è possibile esprimere la correlazione tra un set di variabili osservate in un *path diagram* come somma dei percorsi. È ovviamente possibile invertire questo processo e conoscere i valori dei *path coefficients* a partire dai valori di (inter)correlazione tra le variabili ottenuti da un test empirico. I valori di intercorrelazione rappresentano infatti i valori di relazione tra una variabile ed un'altra, che si otterrebbero dalla somma dei possibili percorsi. Mettendo a sistema tutte le equazioni che esprimono le relazioni tra variabili – se il sistema ammette soluzioni – possiamo ottenere i valori di ogni percorso.

Esempio 3

Supponendo di analizzare il diagramma in Figura A3 (sinistra), semplificazione della Figura A1, possiamo ora scrivere le equazioni che mettono in relazione i percorsi e i valori di intercorrelazione del test empirico:

$$\begin{aligned} r_{AB} &= c \\ r_{AC} &= a + cd \\ r_{BC} &= b + ca \end{aligned}$$

Supponendo di aver ottenuto da un test empirico i valori di intercorrelazione in Figura 4 (destra), possiamo risolvere le equazioni r_{AB} , r_{AC} e r_{BC} :

$$\begin{array}{lcl}
 c = 0,50 & & \\
 a + cb = 0,65 & \Rightarrow & a + 0,50 b = 0,65 \Rightarrow a = 0,40 \\
 b + ca = 0,70 & & 0,50 a + b = 0,7 \Rightarrow b = 0,50
 \end{array}$$

Più nel dettaglio i risultati sono definiti *coefficienti parziali di regressione* (*standardized partial regression coefficients*), essendo dei risultati di regressione indicano una misura dell'influenza di una variabile sull'altra, ed essendo parziali indicano il valore d'influenza quando tutte le altre variabili rimangono costanti.

Se il nostro modello sarà corretto (si veda paragrafo sul test d'ipotesi II.III), potremmo concludere che le influenze descritte dal modello saranno rappresentate dai valori ottenuti (*path coefficients*) per i percorsi ipotizzati.

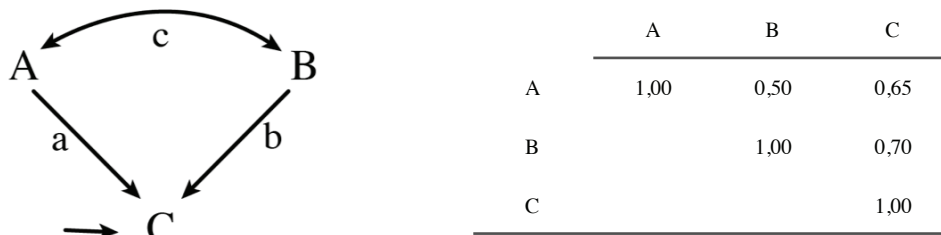


Figura A3. *Path diagram* e valori di intercorrelazione osservati per A, B e C.

I.IV Equazioni strutturali

Un modo alternativo di rappresentare il *path diagram* è quello delle equazioni strutturali. È un sistema composto da tante equazioni quante sono le variabili endogene del sistema e dove ogni equazione descrive le relazioni dirette delle variabili indipendenti sulla variabile dipendente. Ogni variabile dipendente (endogena) è posta uguale alla somma dei *path coefficients* moltiplicati per il valore della variabile indipendente (endogena o esogena) corrispondente, più l'errore stocastico (e) associato alla variabile dipendente. Le variabili esprimono quali sono le influenze sulla variabile dipendente e i coefficienti indicano la “forza di influenza”. A differenza del *path diagram*, nelle equazioni le correlazioni tra le variabili (freccie curve) non sono esplicitate.

Esempio 4

In riferimento al diagramma in Figura A1, l'equazione strutturale per l'unica variabile endogena C è:

$$C = aA + bB + dX + e$$

La Figura A2(a) è invece composta da un modello con tre equazioni per tre variabili endogene:

$$A = aC + e$$

$$F = eC + e$$

$$C = bD + dE + e$$

È importante tener presente che un modello ad equazioni strutturali è un “modello stocastico nel quale ogni equazione rappresenta un legame causale, piuttosto che una mera associazione empirica” (Goldberger 1972). Significa che l'equazione descritta nel modello non ha senso solo da un punto di vista statistico (esprimendo una relazione esistente) ma ha anche carattere causale, ovvero che oltre a “predire” la variabile dipendente la “spiega”. Il modello di regressione semplice non avanza ipotesi sui nessi casuali tra variabili e non permette di cogliere gli effetti indiretti o spuri; il modello ad equazioni strutturali permette invece di tenere conto delle cause che agiscono su una variabile dipendente (analisi multivariata), ma anche delle connessioni (dirette e indirette) esistenti tra le cause.

I.V Equazioni a più soluzioni

Intuitivamente, è facile capire che un sistema di equazioni – in alcuni casi – può ammettere diverse soluzioni. In questi casi, per definire quali valori accettare, si applica il metodo della massima verosimiglianza¹⁰⁰, i risultati devono essere quelli che permettono di minimizzare la differenza tra il modello teorico e i dati osservati. A partire dai dati e dal modello proposto, la stima dei parametri avviene identificando quelli che generano lo scarto minore fra la matrice di covarianza prodotta dal modello e la matrice di covarianza osservata nei dati. Il valore dello

¹⁰⁰ Principio di verosimiglianza: “tra tutti i possibili valori del parametro θ , si preferisce quello che corrisponde alla massima probabilità di generare i dati osservati” (Piccolo, 2010, p.462).

scarto e poi al centro del test di ipotesi del chi-quadro, che valida la corrispondenza tra i dati e le relazioni ipotizzate.

Esempio 5

Il calcolo dei coefficienti avviene in forma matriciale ma – a titolo esemplificativo – può essere mostrato come soluzioni interattive di risoluzione del sistema di equazioni. Dato il diagramma in Figura A4 si inizia il ciclo assegnando dei valori di prova arbitrati di a e b , ad esempio 0,5. A partire da questi si calcolano i valori di correlazione r_{AB} , r_{AC} e r_{BC} implicati dal diagramma, che sono rispettivamente 0,50; 0,50 e 0,25. Definito un criterio di stima della discrepanza tra le correlazioni osservate e ipotizzate (in questo caso la somma delle differenze al quadrato, $\sum d^2$), il residuo è calcolato da $0,11^2 + (-0,08)^2 + (-0,02)^2 = 0,0189$.

Negli *step* successivi si procede per tentativi, modificando i valori di correlazione e stimando i residui. Il processo termina quando si è ottenuto il valore del residuo minimo. In questo caso i valori stimati di correlazione sono $r_{AB}=0,603$, $r_{AC}=0,409$ ed $r_{BC}=0,246$, tale che: $a=0,603$; $b=0,409$; $\sum d^2 = 0,0004480$.

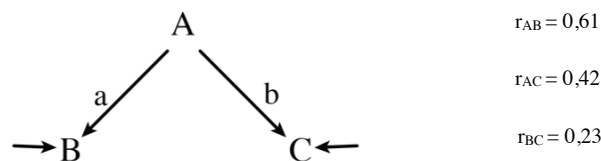


Figura A4. *Path diagram* esempio 5.

II Le fasi del LISREL

Come anticipato, il punto di partenza di un'analisi LISREL è un campionamento, ed in particolare lo studio della matrice di varianza o covarianza tra le variabili osservate. I punti di arrivo sono:

- la definizione dei parametri del modello di equazioni strutturali (i *path coefficients*) che descrivono i nessi casuali tra le variabili.
- il test di ipotesi (basato sul chi-quadro) che valida il modello proposto.

La logica del test di ipotesi si basa sul fatto che se una matrice di covarianza è compatibile con diversi modelli causali tra le variabili, non è vero il contrario, per cui un modello causale può produrre solo una matrice di covarianza.

Il test del modello LISREL si fonda perciò sulla verifica che il modello causale proposto possa essere generato dalla matrice di covarianza dei dati osservati. I

risultati del test di ipotesi confermano la validità delle relazioni causali ipotizzate, ma non escludono l'esistenza di altri modelli compatibili con i dati.

A questo punto può essere utile introdurre una tipica differenziazione dei modelli LISREL, cioè tra analisi esplorativa confermativa.

Nella prima il ricercatore non ha informazioni su quali possono essere le possibili relazioni causali tra le variabili, in questo caso si tenta di costruire il modello, verificarne la validità, e testare modifiche allo stesso. Nell'analisi confermativa invece, il ricercatore conosce già le relazioni tra le variabili, il test è perciò adottato principalmente per conoscere i valori di intensità delle relazioni ovvero i *path coefficients*.

II.I Dati grezzi o dati standardizzati

Un piccolo *excursus* su sul tipo di dati utilizzati nell'analisi è necessario. Infatti le analisi LISREL possono essere effettuate con dati standardizzati o grezzi. Nel primo caso si analizza la correlazione, nel secondo la covarianza. Nel primo caso una freccia dritta in un *path diagram* indica un coefficiente parziale regressione standardizzato, nel secondo caso indica il punteggio grezzo. I coefficienti di regressione standardizzati sono utili nella comparazione tra differenti variabili, mentre i dati grezzi quando si confrontano popolazioni differenti. Inoltre, un punto abbastanza ovvio, è che se si confrontano gruppi diversi e si vuole prendere in considerazione la differenza di varianza, si dovrebbe analizzare la matrice di covarianza; d'altra parte, se si vuole ignorare la varianza, si può scegliere di analizzare le variabili standardizzate, cioè la matrice di correlazione.

II.III Adattamento del modello e test del χ^2

Come anticipato, il test di non falsificazione (o incongruità) del modello con i dati a disposizione, parte dall'analisi della matrice di covarianza tra le variabili osservate X e Y, che è chiamata S. Il modello teorico produce una seconda matrice di covarianza fra le variabili X e Y che abbiamo chiamato C. Il test di falsificazione è basato sul residuo, cioè lo scarto S - C. Nel modello LISREL questo test è costituito dal test di ipotesi del chi quadro (χ^2).

LISREL calcola il valore del χ^2 che confrontato con le tavole di distribuzione ad una probabilità $P = 0,05$ che indica il rifiuto o il non-rifiuto del modello.

Esempio 6

Ipotizziamo di avere 6 valori indipendenti in S e 4 incognite (*path coefficients*) da stimare ($6-4=2$ gradi di libertà df), ed un valore di $\chi^2 = 10,8$; in accordo con le tavole del χ^2 con 2 df è richiesto un valore di 5,99 per rigettare l'ipotesi nulla ad un livello di significatività di 0,05. In questo caso, con il valore di $\chi^2 = 10,8 > 5,99$, si deve rifiutare l'ipotesi nulla e concludere che il modello non si adatta ai dati osservati. Cioè, che la differenza tra C ed S è troppo grande per essere causata da errore campionario. È importante tenere a mente che il test del chi-quadro consente di concludere che un particolare modello non si adatta ai dati ed è quindi errato, ma nel caso in cui χ^2 fosse minore del valore soglia non si può concludere il contrario – che il modello è corretto – ma solo che il test non ha mostrato che questo è errato. Allo stesso modo non si può concludere che tra un modello significativo ed uno non significativo ci sia una differenza statisticamente significativa. Comparazioni tra modelli possono essere fatte solo a determinate condizioni, ad esempio quando i modelli hanno una relazione nidificata o gerarchica (per approfondimenti, non necessari in questa sede, si veda Loehlin 2004).

II.IV Problemi del test χ^2 , l'indice di adattamento RMSEA

Il test del chi quadro è noto per essere sensibile alla numerosità campionaria, con campioni ampi, una discrepanza ridotta può portare al rifiuto del modello quando questo è vero, con campioni piccoli può essere non significativa anche nel caso di misure che non si adattano (*gross misfits*). Per questo una misura più popolarmente utilizzata, relativamente insensibile al numero di unità statistiche del campione, è l'indice di adattabilità *Root Mean Square Error of Approximation* $RMSEA = \sqrt{(d/df)}$. Dove d è ottenuto dividendo $(\chi^2 - df)$ per $(N - 1)$. Per interpretare il valore di RMSEA il valore 0,05 o meno è comunemente indicato come un *fit* del modello “molto buono”, ed un valore 0,10 o inferiore “buono”. (si veda Loehlin 2004, p.69)

Esempio 7

Ipotizzando di aver ottenuto un valore di $\chi^2 = 9,25$ con 2 df da un campione di 408 unità statistiche, il parametro d è $(9,25 - 2) / 407 = 0,018$, perciò RMSEA sarà $= \sqrt{(0,018/2)} = 0,094$.

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

In questo caso il modello è ritenuto “buono” perché compreso tra 0,05 e 0,10.

II.V Miglioramento del modello, l’esclusione dei parametri

Ottenuto un modello significativo è possibile migliorarlo andando a valutare alcuni aspetti di significatività dei singoli parametri per poi migliorare l’adattamento complessivo del modello.

La significatività dei singoli parametri è verificata – nell’ipotesi che la distribuzione campionaria sia normale – dal valore stimato *p-value*, che, se in valore assoluto è maggiore di 1,96 SE (*Standard Error*), indica che l’ipotesi nulla debba essere respinta, cioè il valore del parametro non è significativo; è perciò necessario eliminare il percorso che non ha valenza statistica. Il miglioramento del modello avviene anche attraverso l’inclusione di nuovi parametri o attraverso la riformulazione del modello. Queste procedure non sono trattate in questo lavoro e perciò non saranno richiamate. È sufficiente capire intuitivamente che nel primo caso si tratta di un’aggiunta dei percorsi che spiegano le connessioni tra le variabili, mentre nel secondo caso si tratta di un’aggiunta di variabili latenti che spiegano diversamente il modello.

Appendice B: Il *rough set approach*

Tra gli strumenti dell'analisi di dati incerti, vaghi o inesatti le teorie del *rough set* e del *fuzzy set* sono le due alternative alla teoria degli insiemi ordinaria. I due approcci si differenziano per due distinti metodi di “cattura” degli aspetti della conoscenza imperfetta: l'indiscernibilità nel *rough set* (che per semplicità può essere intesa come uguaglianza) e la vaghezza nel *fuzzy set*. Il primo approccio fu introdotto da Pawlak nel 1985, il secondo da Zadeh nel 1965.

Ciò che differenzia il *rough set* dagli altri approcci della teoria degli insiemi è l'uso di “insiemi non nitidi”. Nella teoria ordinaria un insieme è definito da una serie di sottoinsiemi di elementi, la funzione di appartenenza, che descrive l'appartenenza degli elementi dell'universo ad un insieme, può essere definita solo da due valori: 0 e 1, appartiene o non appartiene. Questo significa che ogni elemento è all'interno o all'esterno dell'insieme in esame. Significa che la definizione della funzione di appartenenza non tiene conto dell'incertezza d'appartenenza di un oggetto di un più ampio insieme di elementi.

Inizialmente, per affrontare i problemi di incertezza è stato introdotto l'approccio *fuzzy*. Questo è definito dalla funzione di appartenenza che può raggiungere valori dell'intervallo 0 e 1, consentendo l'adesione parziale di un oggetto ad un elemento.

Mentre nella *rough set theory*, l'appartenenza non è il concetto primario. Gli insiemi *rough* rappresentano un approccio matematico alla vaghezza e incertezza differente; la definizione di un insieme è relativa alle nostre conoscenze ed alla percezione degli elementi nell'universo. La metodologia del *rough set* si basa sulla proposizione logica che la riduzione del livello di descrizione dei dati rende il modello dei dati più visibile (set elementare), mentre la premessa filosofica è che la conoscenza consiste nella capacità di classificazione. In pratica significa evidenziare i gruppi elementari per cui gli oggetti sono indiscernibili e definire con certezza l'appartenenza di quegli elementi ad un gruppo definito dalle informazioni di conoscenza del decisore. In altre parole il *rough set* può essere considerato come un metodo per definire evidenze da dati imperfetti. Il risultato del *rough set approach* si presenta in forma di classificazione o regole di decisione derivate da un insieme di esempi.

I Concetti base della *rough set theory*

I.1 *Information system*

Il sistema delle informazioni *IS* (o tabella delle informazioni o *approximation space*) è rappresentata da una 4-upla $IS = \langle U, Q, V, f \rangle$ dove U è l'universo (un insieme finito di oggetti x) $U = \{x_1, x_2, \dots, x_n\}$, $Q = \{q_1, q_2, \dots, q_m\}$ è il set di attributi m , $V = \bigcup_{q \in Q} V_q$ è l'insieme dei domini degli attributi e $f : U \times Q \rightarrow V$ la funzione di informazione tale che $f(x, q) \in V_q \forall q \in Q$ e $\forall x \in U$. Ogni oggetto x dell'universo U è descritto dal vettore $Des_q(x) = [f(x, q_1), f(x, q_2) \dots f(x, q_m)]$ chiamato descrizione (*description*).

Esempio 1

Considerando un insieme di 10 oggetti x descritti da 3 attributi il risultato può essere organizzato in una matrice (10 \times 3).

2	1	3
3	2	1
2	1	3
2	2	3
1	1	4
1	1	2
3	2	1
1	1	4
2	1	3
3	2	1

In accordo con la terminologia della *rough sets theory*, questi dati possono essere considerati come una tabella delle informazioni *IS*, dove l'insieme U è costituito dai 10 oggetti $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$ e l'insieme Q dai 3 attributi $Q = \{q_1, q_2, q_3\}$. Mentre gli insiemi V_q , cioè i possibili valori di q sono: $V_{q1} = \{1, 2, 3\}$ $V_{q2} = \{1, 2, 4\}$ e $V_{q3} = \{1, 2, 3, 4\}$.

Il sistema di informazione, per i valori esempio proposti, è presentato in Tabella B1.

Tabella B1. Sistema di informazione IS .

U	q ₁	q ₂	q ₃
x ₁	2	1	3
x ₂	3	2	1
x ₃	2	1	3
x ₄	2	2	3
x ₅	1	1	4
x ₆	1	1	2
x ₇	3	2	1
x ₈	1	1	4
x ₉	2	1	3
x ₁₀	3	2	1

I.II Relazione di indiscernibilità

Per ogni set di attributi $P \subset Q$ è definita una relazione di indiscernibilità (*indiscernibility relation*) $Ind(P)$ o I_P quando due oggetti x_i e x_j sono indiscernibili dal set di attributi di P contenuto in Q se $q(x_i) = q(x_j)$ per ogni $x \in P$. La relazione di indiscernibilità può essere intesa per l'intero insieme $P \subset Q$ o per un suo sottoinsieme $O \subset P \subset Q$.

In altre parole $I_P = \{(x,y) \in U \times U : f(x,q) = f(y,q), q \in P\}$ dove P indica un qualunque sottoinsieme di Q . In termini matematici questa relazione non è altro che una relazione di equivalenza; due elementi Q_1 e Q_2 sono indiscernibili quando tutti i loro attributi sono equivalenti $(x_{11}, x_{12}, x_{13}) = (x_{21}, x_{22}, x_{23})$.

Esempio 2

Nel sistema di informazione proposto, alcuni degli oggetti sono identici, come ad esempio x_1 e x_3 ; questi sono indiscernibili. Raggruppando gli elementi per indiscernibilità, considerando tutte e tre le variabili q_x si ottengono 5 gruppi (Tabella B2).

Tabella B2. Gruppi di indiscernibilità nel set P .

U / P	q ₁	q ₂	q ₃
{x ₁ , x ₃ , x ₉ }	2	1	3
{x ₂ , x ₇ , x ₁₀ }	3	2	1
{x ₄ }	2	2	3
{x ₅ , x ₈ }	1	1	4
{x ₆ }	1	1	2

Ogni riga descrive un set elementare (o granulo di conoscenza) che rappresenta il più piccolo gruppo indiscernibile degli oggetti del set $P=\{q_1, q_2, q_3\}$.

Quando la relazione di indiscernibilità è limitata ad alcune variabili, per esempio del sottoinsieme $O=\{q_1, q_2\}$, gli insiemi si limitano a considerare la relazione di equivalenza delle due variabili, ignorando gli altri dati (Tabella B3).

Tabella B3. Gruppi di indiscernibilità nel set O .

U / O	q_1	q_2
$\{x_1, x_3, x_9\}$	2	1
$\{x_2, x_7, x_{10}\}$	3	2
$\{x_4\}$	2	2
$\{x_5, x_6, x_8\}$	1	1

I.III Approssimazione inferiore e superiore

L'approccio *rough sets* dipende da due concetti base, l'approssimazione inferiore (*lower approximation*) e l'approssimazione superiore (*upper approximation*) dei set (Figura B1).

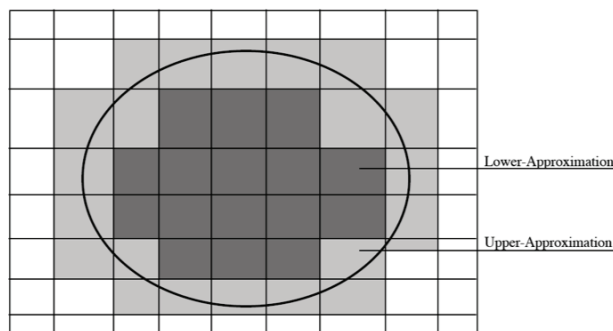


Figura B1. Rappresentazione schematica dell'approssimazione inferiore e superiore.

L'approssimazione inferiore di P si compone di tutti e solo quegli elementi che appartengono indubbiamente all'insieme, cioè che appartengono alle classi di equivalenza generate dalla relazione di indiscernibilità contenute in X . Dove X è un sottoinsieme dell'universo U ($X \subset U$).

L'approssimazione superiore di P si compone di tutti quegli elementi che possono appartenere al set, cioè che appartengono alle classi di equivalenza che contengono almeno un oggetto appartenente ad X .

Dato il sottoinsieme X dell'universo U , le approssimazioni inferiori e superiori sono definite da:

$$\underline{P}(X) = \{x \in U : I_p(x) \subseteq X\}$$

$$\bar{P}(X) = \{x \in U : I_p(x) \cap X \neq \emptyset\}$$

Dove:

P -lower del set X è il set degli oggetti x_i , che appartengono al set elementare nello spazio di X .

P -upper del set X è l'unione dei set elementari, che non hanno intersezione vuota con X .

Ogni oggetto x_i dell'approssimazione P -lower di X ($x_i \in \underline{P}(X)$) è certo che appartenga ad X , mentre per ogni oggetto x_i dell'approssimazione P -upper di X ($x_i \in \bar{P}(X)$) possiamo solo sapere che può appartenere ad X .

La differenza è chiamata *boundary* di X in U e sta ad indicare gli elementi che non possono essere classificati con certezza in X usando le informazioni disponibili.

Formalmente:

$$BN_p(X) = \bar{P}(X) - \underline{P}(X)$$

Quando per l'insieme X non esistono oggetti nella frontiera, cioè $\underline{P} = \bar{P} \Rightarrow BN = \emptyset$ allora l'insieme X è definito ordinario (esatto) e può essere descritto in maniera certa come unione di classi di equivalenza. Quando invece $\underline{P} \neq \bar{P} \Rightarrow BN \neq \emptyset$ l'insieme X è definito come insieme approssimato.

Ci sono quattro possibili tipi di set indefiniti, dove \emptyset indica l'insieme vuoto:

1. Se $\underline{P}(x) \neq \emptyset$ AND $\bar{P}(x) \neq U$, X è chiamato *roughly definable* in U
2. Se $\underline{P}(x) \neq \emptyset$ AND $\bar{P}(x) = U$, X è chiamato *externally undefinable* in U
3. Se $\underline{P}(x) = \emptyset$ AND $\bar{P}(x) \neq U$, X è chiamato *internally undefinable* in U
4. Se $\underline{P}(x) = \emptyset$ AND $\bar{P}(x) = U$, X è chiamato *totally undefinable* in U

Esempio 3

Supponendo di essere interessati al *subset* X composto da cinque oggetti $X = \{x_1, x_3, x_4, x_5, x_9\}$, possiamo distinguere questa serie dall'intero insieme di dati nello spazio di tre attributi $P = \{q_1, q_2, q_3\}$. Per le approssimazioni inferiori, i set

elementari contenuti in X sono: $\{x_1, x_3, x_9\}, \{x_4\}$. L'approssimazione superiore sarà perciò: $\underline{P} = \{x_1, x_3, x_4, x_9\}$. L'approssimazione inferiore è invece individuata a partire dai set elementari che hanno almeno un elemento in comune con il subset X , questi sono: $\{x_1, x_3, x_9\}, \{x_4\}, \{x_5, x_8\}$. L'approssimazione superiore è perciò: $\bar{P} = \{x_1, x_3, x_4, x_5, x_8, x_9\}$.

Di conseguenza il confine BN contiene gli elementi che sono in P -upper ma non in P -lower. $BN = \{x_1, x_3, x_4, x_5, x_8, x_9\} - \{x_1, x_3, x_4, x_9\} = \{x_5, x_8\}$

L.IV Accuratezza e qualità dell'approssimazione

Alcuni indici sintetici della accuratezza delle misure nel set X sono definite come:

$$\alpha_p(X) = \frac{|\underline{P}(X)|}{|\bar{P}(X)|}$$

$$\gamma_p(X) = \frac{|\underline{P}(X)|}{|X|}$$

$$\mu_X^P(x) = \frac{|X \cap I_p(x)|}{|I_p(x)|}$$

dove α_p , ottenuta dal rapporto tra P -lower e P -upper, è l'indice di accuratezza (*accuracy*) dell'approssimazione, cioè il rapporto tra gli oggetti assegnati con certezza all'insieme X e gli oggetti che possono appartenere all'insieme X . L'indice di accuratezza può assumere valori $0 \leq \alpha_p \leq 1$ (compresi tra 0 ed 1), se X è definibile in U allora $\alpha_p = 1$, se è indefinibile in U allora $\alpha_p < 1$.

La qualità dell'approssimazione γ_p (*quality of approximation*) è invece ottenuta dal rapporto tra gli oggetti assegnati con certezza all'insieme X e l'universo degli oggetti di X . μ_X^P rappresenta la funzione di appartenenza (*rough membership function*) di un oggetto x ed indica la credibilità (*degree of certainty*) della appartenenza al sottoinsieme X .

Esempio 4

In riferimento ai dati del sottoinsieme $X = \{x_1, x_3, x_4, x_5, x_9\}$, il numero di oggetti contenuti in P -lower è 4, mentre in P -upper 6. L'accuratezza dell'approssimazione è $\alpha_p = 4/6$. La qualità dell'approssimazione è invece ottenuta da 4 elementi contenuti in P -lower fratto 10 oggetti dell'universo: $\gamma_p = 4/10$.

I.V Indipendenza degli attributi

Può capitare che uno degli attributi non presenti caratteristiche che lo rendano indipendente, cioè indispensabile alla misurazione. Per accertarsi che gli attributi siano indipendenti, si verifica per ogni attributo che la sua assenza modifichi il numero di set elementari per gli oggetti esaminati.

Formalmente se $I_Q = I_{Q-q_i}$ allora l'attributo q_i è superfluo, altrimenti q_i è indispensabile in Q .

Esempio 5

Considerando i dati in Tabella B1, 3 attributi in Q (q_1 , q_2 e q_3) e 10 oggetti in U , possono essere costruiti 5 set elementari (Tabella B2). La Tabella B4 restituisce in numero di set elementari dopo aver eliminato uno degli attributi. Ad esempio, rimuovendo q_1 il numero di set elementari non cambia, rimuovendo invece q_2 o q_3 gli insiemi si riducono a 4. Questo significa che q_1 è indipendente, cioè superfluo.

Tabella B4. Numero di set elementari rimuovendo un attributo.

	Attributi rimossi			
	Nessuno	q_1	q_2	q_3
Numero di set elementari	5	5	4	4

I.VI Nucleo e sottoinsiemi ridotti

Se il set di attributi è dipendente, si può essere interessati a individuare tutti i possibili sottoinsiemi ridotti (*reduct*) degli attributi $P' \subseteq P$ che permettono di mantenere inalterata la qualità dell'approssimazione $\gamma_{P'} = \gamma_P$ e di individuare il set di tutti gli attributi indispensabili, cioè il nucleo (*core*) che corrisponde all'intersezione di tutti gli insiemi ridotti. *Reduct* e *core* sono due concetti fondamentali della *rough sets theory*, il primo costituisce la parte essenziale di un *IS*, il secondo è la parte comune di ogni *reducts*. La loro misura è ottenuta dalla matrice di discernibilità (*discernibility matrix*) che ha una dimensione $n \times n$, dove n indica il numero di set elementari. La matrice di discernibilità D è costruita dal calcolo degli elementi d_{ij} , ottenuti dagli elementi non in comune tra i set base i e j .

Esempio 6

Considerando i dati in Tabella B2, il set di attributi da cui discerne il se 1 e 2 contiene gli attributi q_1 , q_2 e q_3 , perciò $d_{21} = d_{12} = \{q_1, q_2, q_3\}$. Gli elementi d_{12} e

d_{21} sono composti dagli attributi che presentano valori diversi, infatti $q_{11} = (2) \neq q_{21} = (3)$, $q_{12} = (1) \neq q_{22} = (2)$ e $q_{13} = (3) \neq q_{23} = (1)$. d_{31} è invece composta da un unico valore q_2 , in quanto $q_{11} = q_{31}$, $q_{12} \neq q_{22}$ e $q_{32} = q_{33}$.

La diagonale inferiore della matrice simmetrica ($d_{ij} = d_{ji}$) per i dati in Tabella B2 è riportata in Tabella B5. Il valore di *core* è calcolato come l'insieme di tutti gli elementi singoli della matrice di discernibilità. Mentre il valore di *reduct* è il sottoinsieme degli attributi che hanno almeno un elemento in comune con ogni voce non vuota nella matrice. Nel caso esempio *core* e *reduct* coincidono e sono definiti da: $q_2 q_3$.

Tabella B5. Matrice di discernibilità.

	Set 1	Set 2	Set3	Set 4	Set 5
Set 1					
Set 2	q_1, q_2, q_3				
Set 3	q_2	q_1, q_3			
Set 4	q_1, q_3	q_1, q_2, q_3	q_1, q_2, q_3		
Set 5	q_1, q_3	q_1, q_2, q_3	q_1, q_2, q_3	q_3	

I.VII Decision table

Se all'interno di una tabella delle informazioni $IS = \langle U, Q, V, f \rangle$ si distinguessero gli attributi Q in attributi condizionali $C \neq \emptyset$ ed uno o più attributi decisionale $D \neq \emptyset$, tale che $C \cup D = Q$ e $C \cap D = \emptyset$, si otterrebbe una tabella decisionale S (*decision table*). L'attributo decisionale impone una partizione dell'universo U tramite relazioni di indiscernibilità I_D in classi di decisione (*decision classes*) indipendenti dagli attributi condizionali di C . Questo tipo di tabelle decisionali possono essere usate per capire se le proprietà di un oggetto espresse in termini di attributi condizionali C siano descrivibili in termini di attributi decisionali D . Questo permette, ad esempio, di comparare le preferenze (soggettive) del decisore in termini di indicazioni condizionali (oggettive).

Esempio 7

A partire dall'insieme di 10 oggetti descritti da 3 attributi condizionali in Tabella B1, integrando le preferenze di tre decisori $D = \{d, d', d''\}$ con valori $V_d = \{1, 2, 3\}$ si può costruire la tabella decisionale in Tabella B6.

In accordo con le scelte del decisore d l'appartenenza ad ogni classe degli oggetti x è:

Classe 1: $\{x_1, x_3, x_9\}$

Classe 2: $\{x_2, x_4, x_7, x_{10}\}$

Classe 3: $\{x_5, x_6, x_8\}$

Tabella B6. Tabella decisionale (*decision table*).

U	q ₁	q ₂	q ₃	d	d'	d''
x ₁	2	1	3	1	2	3
x ₂	3	2	1	2	3	1
x ₃	2	1	3	1	2	3
x ₄	2	2	3	2	3	1
x ₅	1	1	4	3	1	2
x ₆	1	1	2	3	1	3
x ₇	3	2	1	2	3	1
x ₈	1	1	4	3	1	3
x ₉	2	1	3	1	2	3
x ₁₀	3	2	1	2	3	1

I.VIII Regole di decisione

Lo scopo della procedura è quella di evidenziare le relazioni di dipendenza tra attributi condizionali e decisionali, restituendo alcune regole decisionali (*decision rules*) del tipo *if...then*, formalmente:

If $f(x, c_1) = r_{c1}$ and ... and $f(x, c_P) = r_{cP}$ then $x \in D_1$ or ... or $x \in D_k$

In altre parole $q_{k_i} \Rightarrow d_j$ dove q_{k_i} indica che "l'attributo q_k ha valore i " e il simbolo " \Rightarrow " indica una implicazione posizionale (allora, *then*). In generale la regola di decisione $\Theta \Rightarrow \Phi$, è composta da Θ condizione (*condition*) e Φ decisione (*decision*). La minimizzazione degli insiemi di attributi e valori (derivata dalla matrice di indiscernibilità) è semplicemente una riduzione delle condizioni inutili alla definizione delle regole di decisione. Le regole sono infatti derivate considerando gli attributi condizionali che non sono indipendenti (vedi indipendenza degli attributi). Significa individuare i granuli di conoscenza che supportano le regole.

Esempio 8

Considerando la porzione della tabella decisionale (Tabella B6) composta dal set di elementi $X=\{x_1, x_2, \dots, x_{10}\}$, gli attributi $Q=\{q_1, q_2\}$ e il set decisionale $D=\{d\}$, in Tabella B7 sono elencati i valori che definiscono le regole di decisione:

$$q_{1_2} \text{ AND } q_{2_1} \Rightarrow d_1$$

$$q_{2_2} \Rightarrow d_2$$

$$q_{1_1} \Rightarrow d_3$$

dove d descrive il set elementare in D .

La Tabella B8 mostra un'esemplificazione grafica dei granuli di conoscenza che definiscono le regole.

Tabella B7. Sistema di informazione IS .

U	q ₁	q ₂	d
x ₁	2	1	1
x ₃	2	1	1
x ₉	2	1	1
x ₄	(2)*	2	2
x ₂	(3)*	2	2
x ₇	(3)*	2	2
x ₁₀	(3)*	2	2
x ₅	1	(1)*	3
x ₆	1	(1)*	3
x ₈	1	(1)*	3

* indica "non è importante", ovvero che compongono regole secondarie

Tabella B8. Esempificazione dei granuli di conoscenza per i sottoinsiemi.

Attributo condizionale	q ₁	1			2			3			
Attributo decisionale	d	3	3	3	1	1	1	2	2	2	2

Attributo condizionale	q ₂	1			2						
Attributo decisionale	d	1	1	1	3	3	3	2	2	2	2

Attributo condizionale	q ₁	1			2			3			
	q ₂	1			2			2			
Attributo decisionale	d	3	3	3	1	1	1	2	2	2	2

L.IX Regole certe, possibili o approssimate

Le regole decisionali possono essere determinate in tre modi: dall'approssimazione inferiore *P-lower*, da quella superiore *P-upper* o dalla frontiera *BN*. Cioè il sottoinsieme di oggetti che definiscono la regola appartengono ad uno ed uno solo dei tre possibili set.

Se la regola risulta determinata da un'approssimazione inferiore di una classe di decisione, si può assegnare con certezza un dato oggetto ad una univoca classe di decisione, in questo caso la regola è definita certa (*certain decision rules*).

Quando la regola è determinata dall'approssimazione superiore di una classe decisionale, non si può essere certi dell'assegnazione di un oggetto ad una determinata classe, si parla di regola possibile (*possible decision rules*).

Se invece la regola deriva dalla frontiera allora si definisce approssimata o ambigua in quanto un oggetto può essere assegnato a diverse classi di decisione.

Può capitare che alcune regole siano contraddittorie o incompatibili, per definire un valore di importanza (e quindi di priorità) tra le regole sono introdotti tre indici.

Il supporto (*support*) ovvero il numero di oggetti $x \in U$ che soddisfano la parte condizionale e decisionale della regola, ovvero il numero di oggetti che definiscono e supportano la regola.

Il grado di copertura della regola (*cover*) è invece il rapporto tra il numero di oggetti che soddisfano sia la parte condizionale che quella decisionale e gli oggetti che soddisfano solo la parte condizionale. È quindi il rapporto tra il valore *support* ed il numero di oggetti simili. Cioè la capacità della relazione di spiegare i fenomeni osservati.

La forza della regola (*strength*) è il rapporto tra gli oggetti che soddisfano entrambe le parti (condizionale e decisionale) sul totale degli oggetti presenti nell'universo di riferimento. Cioè il rapporto tra il valore *support* ed il numero totale degli oggetti appartenenti con certezza alla classe. Il valore di *strength* fornisce la rappresentatività della relazione osservata.

Al fine di definire la priorità tra due o più regole il valore di *strength* più alto è quello che definisce la regola principale.

Esempio 9

In accordo con i dati proposti, per le due regole $q_{1_1} \Rightarrow d_3$ e $q_{2_2} \Rightarrow d_2$ il numero di oggetti che soddisfano sia la parte condizionale che decisionale (*support*) sono rispettivamente 3 (x_5, x_6, x_8) e 4 (x_2, x_4, x_7, x_{10}).

Considerando che le $x \in P_3$ sono 3 e le $x \in P_2$ sono 7 i valori di forza della regola sono rispettivamente $3/3=100\%$ e $4/7=57\%$.

I.X Nuove decisioni

Le regole di decisione derivate dai dati ottenuti in maniera sperimentale possono essere adottate per supportare nuove decisioni e classificazioni. La corrispondenza di un nuovo elemento ad una classe è derivata confrontando gli attributi dell'oggetto con le regole logiche ottenute. Questa procedura può essere caratterizzata da quattro situazioni:

- (1) Il nuovo oggetto corrisponde esattamente ad una delle regole deterministiche.
- (2) Il nuovo oggetto corrisponde esattamente ad una delle regole non deterministiche.
- (3) Il nuovo oggetto non corrisponde a nessuna regola.
- (4) Il nuovo oggetto corrisponde a più di una regola deterministica.

Nel primo caso (1) l'assegnazione è esatta, l'oggetto è assegnato con certezza ad una classe in accordo con le regole che lo descrivono. Nel caso successivo (2), le regole sono ambigue. In questo caso il decisore conosce l'informazione sul numero di esempi (*support*) che supportano ogni regola e sulla loro forza relativa (*strength*). Se il valore di forza di una regola è migliore del valore delle altre si può concludere che secondo questa regola l'oggetto considerato molto probabilmente appartiene alla classe definita dalla regola più forte. La terza situazione (3) è invece la più complessa, in questo caso si cercano le regole più "vicine" alla descrizione dell'oggetto. La notazione "vicine" comporta l'uso di una misura della distanza. Il nuovo oggetto può essere identificato come valore *outlier* di classi esistenti o può essere considerato come un elemento di una nuova classe. Nell'ultimo caso (4) se tutte le decisioni concordano sull'assegnazione di una stessa classe all'oggetto non c'è ambiguità. In caso contrario si definisce la regola principale in accordo al valore di forza relativa.

Esempio 10

Date le regole $q_{1_2} \wedge q_{2_1} \Rightarrow d_1$, $q_{2_2} \Rightarrow d_2$ e $q_{1_1} \Rightarrow d_3$ ed in aggiunta la regola $q_{1_2} \Rightarrow d_1$ or d_2 si ipotizzi di voler classificare i nuovi oggetti $x^{n_1} = \{1, 3\}$, $x^{n_2} = \{2, 3\}$, $x^{n_3} = \{3, 3\}$ e $x^{n_4} = \{2, 1\}$.

x^{n_1} corrisponde esattamente alla terza regola deterministica (1), l'oggetto è assegnato alla classe D_3 in accordo con la regola.

x^{n_2} corrisponde esattamente alla quarta regola non deterministica (2), l'oggetto contiene una sola caratteristica descritta dalle regole e può essere assegnato alla classe D_1 o D_2 .

x^{n_3} non corrisponde a nessuna delle regole descritte, l'oggetto può essere classificando adottando la regola più vicina (es. $q_{2_2} \Rightarrow d_2$) o può essere classificato dal decisore e adottato per la costruzione di nuove regole.

x^{n_4} corrisponde invece a due regole, in questo caso in contraddizione (4). L'assegnazione dell'oggetto è definita in funzione della forza della regola, ovvero si classifica x in accordo con la regola che ha il valore *strength* maggiore.

II Concetti base del *Dominance-based rough set approach*

L'approccio alla teoria del *rough set* analizza le caratteristiche degli oggetti in funzione della indiscernibilità tra i valori degli attributi, in alcuni casi è però utile confrontare le relazioni di dominanza, piuttosto che limitarsi alla sola equivalenza. Il *dominance-based rough set approach (DRSA)*, così come definito nella letteratura di riferimento, è un approccio che, introdotto da Greco *et al.* (1997, 1999, 2001) sostituisce la relazione di indiscernibilità con la relazione di dominanza. In base a questo principio, se un oggetto x_a presenta almeno le stesse caratteristiche, relativamente ai criteri condizionali, rispetto ad un oggetto x_b allora dovrà essere classificato almeno nella stessa classe di x_b . Il vantaggio di questo approccio sta nel fatto che la costruzione delle regole di decisione è subordinata all'ordine che sussiste fra i valori degli attributi e le classi assegnate; significa che la relazione d'ordine tra attributi e classi è presa come base del confronto che restituisce regole di appartenenza "almeno o al più" ad una classe. È dunque chiaro che l'aspetto della dominanza permette di ampliare le informazioni desumibili dalla relazione di indiscernibilità.

Quindi, formalmente, il DRSA assume che le preferenze aumentino di un valore di $f(\cdot, q)$ per ogni $q \in C$ e che, considerando un unico attributo decisionale $D=\{d\}$ il valore di d sia definito da un numero finito di classi decisionali $Cl=\{Cl_t, t \in T\}$,

$T=\{0, \dots, n\}$, tale che ogni $x \in U$ appartenga ad una ed una sola classe. Inoltre, si suppone che le classi siano ordinate per preferenza, ovvero che per ogni $r, s \in T$ tale che $r > s$, gli oggetti in classe Cl_r siano preferiti agli oggetti in classe Cl_s .

II.I Relazione di dominanza

Considerando $P \subseteq C$ un sottoinsieme degli attributi condizionali, la relazione di dominanza Δ_p associata con P è definita per ogni coppia di oggetti x e y da:

$$x \Delta_p y \Leftrightarrow f(x, q) \geq f(y, q), \forall q \in P$$

Dove il simbolo \geq può essere sostituito con \leq per criteri ordinati in accordo con una preferenza decrescente. In accordo con la relazione di dominanza ogni oggetto $x \in U$ può essere associato a due set: l'insieme *P-dominating*, definito da $\Delta_p^+(x) = \{y \in U : y \Delta_p x\}$ che contiene gli oggetti che dominano x e l'insieme *P-dominated* $\Delta_p^-(x) = \{y \in U : x \Delta_p y\}$ che contiene gli oggetti dominati da x . Questi insiemi sono usati per approssimare le classi decisionali.

Tabella B9. Sistema di informazione IS.

U	P-dominating set	P-dominated set
x_1	$\Delta_p^+(x_1) = \{x_1, x_3, x_4, x_9\}$	$\Delta_p^-(x_1) = \{x_1, x_3, x_4, x_9\}$
x_2	$\Delta_p^+(x_2) = \{x_2, x_7, x_{10}\}$	$\Delta_p^-(x_2) = \{x_2, x_7, x_{10}\}$
x_3	$\Delta_p^+(x_3) = \{x_1, x_3, x_4, x_9\}$	$\Delta_p^-(x_3) = \{x_1, x_3, x_4, x_9\}$
x_4	$\Delta_p^+(x_4) = \{x_4\}$	$\Delta_p^-(x_4) = \{x_1, x_3, x_4, x_6, x_9\}$
x_5	$\Delta_p^+(x_5) = \{x_5, x_8\}$	$\Delta_p^-(x_5) = \{x_5, x_6, x_8\}$
x_6	$\Delta_p^+(x_6) = \{x_1, x_3, x_4, x_5, x_6, x_8, x_9\}$	$\Delta_p^-(x_6) = \{x_6\}$
x_7	$\Delta_p^+(x_7) = \{x_2, x_7, x_{10}\}$	$\Delta_p^-(x_7) = \{x_2, x_7, x_{10}\}$
x_8	$\Delta_p^+(x_8) = \{x_5, x_8\}$	$\Delta_p^-(x_8) = \{x_5, x_6, x_8\}$
x_9	$\Delta_p^+(x_9) = \{x_1, x_3, x_4, x_9\}$	$\Delta_p^-(x_9) = \{x_1, x_3, x_6, x_9\}$
x_{10}	$\Delta_p^+(x_{10}) = \{x_2, x_7, x_{10}\}$	$\Delta_p^-(x_{10}) = \{x_2, x_7, x_{10}\}$

Esempio 11

In accordo con la tabella decisionale (Tabella B6), considerando l'universo degli oggetti $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$, il set di attributi condizionali $C = \{q_1, q_2, q_3\}$ e il set di attributi decisionali $D = \{d\}$ è facile stabilire che x_1 è dominato dagli oggetti x_3, x_4 e x_9 in quanto $f(x_1, C_j) \leq f(x_k, C_j)$, $k=3, 4, 9$; $j=1, 2, 3$. Mentre, ad

esempio, rispetto ad x_2 non c'è relazione di dominanza, in quanto x_1 è migliore per gli attributi q_1 e q_2 , mentre x_2 è migliore in q_3 .

In Tabella B9 sono riassunti gli insiemi *P-dominating* e *P-dominated* associati all'universo U .

II.II Approximating upward and downward class unions

Dal punto di vista dei criteri decisionali le relazioni risultano simili, gli oggetti sono però aggregati in funzione della classe decisionale.

L'unione degli oggetti y che presentano almeno la stessa classe decisionale dell'oggetto x o una superiore costituiscono la *upward union of classes* $CI_D^{\geq x}$.

Mentre la *downward union of classes* $CI_D^{\leq x}$ è l'insieme degli oggetti y che presentano al più la stessa classe decisionale dell'oggetto x o una inferiore.

Formalmente:

$$CI_D^{\geq x} = \{y \in U : y \Delta x\}$$

$$CI_D^{\leq x} = \{y \in U : x \Delta y\}$$

Nel caso in cui il criterio decisionale ripartisca l'universo U in un insieme finito di classi ordinate allora l'unione delle classi *upward* e *downward* può essere riassunta come:

$$CI_t^{\geq} = \bigcup_{s \geq t} CI_s, \quad CI_t^{\leq} = \bigcup_{s \leq t} CI_s$$

L'unione verso l'alto CI_t^{\geq} include gli oggetti che appartengono almeno alla classe CI_t (cioè alla stessa classe o ad una superiore), mentre l'unione verso il basso CI_t^{\leq} è composta dagli oggetti che appartengono al più a CI_t (cioè stessa classe o classe inferiore). L'asserzione $x \in CI_t^{\leq}$ significa che “ x appartiene almeno alla classe CI_t ” mentre $x \in CI_t^{\geq}$ che “ x appartiene al più alla classe CI_t ”.

Esempio 12

In accordo con la tabella decisionale adottata nell'esempio 11 l'unione delle classi CI_t^{\geq} e CI_t^{\leq} per le tre classi decisionali 1, 2, 3 è descritto in Tabella B10.

Tabella B10. Unione delle classi decisionali *upward* e *downward*.

D	Upward union (almeno Cl _i)	Downward union (al più Cl _i)
Cl ₁	Cl ₁ ⁺ ={U}	Cl ₁ ⁻ ={x ₁ , x ₃ , x ₉ }
Cl ₂	Cl ₂ ⁺ ={x ₂ , x ₄ , x ₅ , x ₆ , x ₇ , x ₈ , x ₁₀ }	Cl ₂ ⁻ ={x ₁ , x ₂ , x ₃ , x ₄ , x ₇ , x ₉ , x ₁₀ }
Cl ₃	Cl ₃ ⁺ ={x ₅ , x ₆ , x ₈ }	Cl ₃ ⁻ ={U}

II.III Approssimazione inferiore e superiore

Nell'approccio basato sulla dominanza, le approssimazioni inferiori e superiori (*lower* e *upper approximations*), considerando $P \subseteq C$, sono misurate per le unioni verso l'alto Cl_i^{\geq} e verso il basso Cl_i^{\leq} . Formalmente:

$$\begin{aligned} \underline{P}(Cl_i^{\geq}) &= \{x \in U : \Delta_P^+(x) \subseteq Cl_i^{\geq}\} \\ \bar{P}(Cl_i^{\geq}) &= \bigcup_{x \in Cl_i^{\geq}} \Delta_P^-(x) = \{x \in U : \Delta_P^-(x) \cap Cl_i^{\geq} \neq \emptyset\} \\ \underline{P}(Cl_i^{\leq}) &= \{x \in U : \Delta_P^-(x) \subseteq Cl_i^{\leq}\} \\ \bar{P}(Cl_i^{\leq}) &= \bigcup_{x \in Cl_i^{\leq}} \Delta_P^+(x) = \{x \in U : \Delta_P^+(x) \cap Cl_i^{\leq} \neq \emptyset\} \end{aligned}$$

L'approssimazione inferiore di P per Cl_i^{\geq} contiene tutti quegli elementi dell'insieme P -dominating assegnati con certezza ad una classe almeno buona come Cl_i ovvero gli elementi di $\Delta_P^+(x) \subseteq Cl_i^{\geq}$.

L'approssimazione superiore di P per Cl_i^{\geq} contiene tutti quegli elementi che possono appartenere alla classe Cl_i , cioè che hanno in comune almeno una porzione tra l'insieme P -dominated e la *upward union* $\Delta_P^-(x) \cap Cl_i^{\geq}$.

Il valore di confine *boundary*, in accordo con a regola base è misurato dalla differenza tra P -upper e P -lower.

$$\begin{aligned} BN_p(Cl_i^{\geq}) &= \bar{P}(Cl_i^{\geq}) - \underline{P}(Cl_i^{\geq}) \\ BN_p(Cl_i^{\leq}) &= \bar{P}(Cl_i^{\leq}) - \underline{P}(Cl_i^{\leq}) \end{aligned}$$

Esempio 13

Le approssimazioni superiori e inferiori e i valori di confine di P per le classi Cl_1 , Cl_2 e Cl_3 definite per i dati in Tabella B6 sono riportati in Tabella B11. Per la classe Cl_2 , in accordo con le definizioni precedenti, $\underline{P}(Cl_2^{\geq}) = \{x \in U : \Delta_P^+(x) \subseteq Cl_2^{\geq}\}$. In base alla Tabella 20 sappiamo che $Cl_2^{\geq} = \{x_2, x_4, x_5, x_6, x_7, x_8, x_{10}\}$, usando i set P -dominating definiti in Tabella 19 $\Delta_P^+(x_2) = \Delta_P^+(x_7) = \Delta_P^+(x_{10}) = \{x_2, x_7, x_{10}\}$, Δ_P^+

$(x_4)=\{x_4\}$, $\Delta_P^+(x_5)=\Delta_P^+(x_8)=\{x_5, x_8\}$ e $\Delta_P^+(x_6)=\{x_1, x_3, x_4, x_5, x_6, x_8, x_9\}$

l'approssimazione inferiore di Cl_2^{\geq} è $\underline{P}(Cl_2^{\geq}) = \{x_2, x_4, x_5, x_7, x_8, x_{10}\}$ (elementi di Δ_P^+ contenuti in Cl_2^{\geq}).

Tabella B11-21. Unione delle approssimazioni superiori, inferiori e della regione limite per le classi decisionali Cl_1 , Cl_2 e Cl_3 .

Lower approximations	Upper approximations	Boundaty
$P(Cl_1^{\leq}) = \emptyset$	$P(Cl_1^{\leq}) = \{x_1, x_3, x_6, x_9\}$	$BN_P(Cl_1^{\leq}) = \{x_1, x_3, x_6, x_9\}$
$P(Cl_2^{\leq}) = \{x_2, x_7, x_{10}\}$	$P(Cl_2^{\leq}) = \{x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{10}\}$	$BN_P(Cl_2^{\leq}) = \{x_1, x_3, x_4, x_6, x_9\}$
$P(Cl_2^{\geq}) = \{x_2, x_4, x_5, x_7, x_8, x_{10}\}$	$P(Cl_2^{\geq}) = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$	$BN_P(Cl_2^{\geq}) = \{x_1, x_3, x_6, x_9\}$
$P(Cl_3^{\geq}) = \{x_5, x_8\}$	$P(Cl_3^{\geq}) = \{x_1, x_3, x_4, x_5, x_6, x_8, x_9\}$	$BN_P(Cl_3^{\geq}) = \{x_1, x_3, x_4, x_6, x_9\}$

II.IV *Quality of classification*

Anche per l'approccio fondato sulle dominanze sono calcolati i valori di accuratezza delle approssimazioni, in particolare la qualità dell'approssimazione γ_P (*quality of approximation*) è ottenuta dal rapporto tra gli oggetti correttamente classificati in P (*P-correctly*) e l'universo degli oggetti del sistema.

$$\gamma_P(Cl) = \frac{|U - ((\bigcup_{t \in T} BN_P(Cl_t^{\geq})) \cup (\bigcup_{t \in T} BN_P(Cl_t^{\leq})))|}{|U|}$$

Il valore di accuratezza (*accuracy*) cioè il rapporto tra gli oggetti assegnati con certezza all'insieme Cl_t e gli oggetti che possono appartenere all'insieme Cl_t è ottenuto dal rapporto tra *P-lower* e *P-upper*. Formalmente:

$$\alpha_p(Cl_t^{\geq}) = \frac{|\underline{P}(Cl_t^{\geq})|}{|\overline{P}(Cl_t^{\geq})|}$$

$$\alpha_p(Cl_t^{\leq}) = \frac{|\underline{P}(Cl_t^{\leq})|}{|\overline{P}(Cl_t^{\leq})|}$$

Esempio 14

Per i dati proposti in esempio, i valori di accuratezza sono descritti in Tabella B12. A titolo di esempio, l'accuratezza $\alpha_p(Cl_2^{\geq}) = 6$ (*lower-approximation*) / 10 (*upper-*

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

approximation) = 0,6; il valore di accuratezza $\alpha_p(CI_l^{\leq}) = 0$ (*lower-approximation*) / 3 (*upper-approximation*) = 0,0.

Il valore di qualità dell'approssimazione è pari a $\gamma_p(CI) = ((U-BN)/U) = ((10-5)/10)=0,5$.

Tabella B12. Valori di accuratezza.

Union	Accuracy	Card Union	Low-appr. Card	Upper-appr. Card
At most 1	0,00	3	0	4
At most 2	0,38	7	3	8
At least 2	0,60	7	6	10
At least 3	0,29	3	2	7

II.V Regole di decisione

A differenza dell'approccio classico dove le regole di decisione definiscono l'appartenenza di un oggetto ad una classe, nell'approccio basato sulle dominanze le regole restituiscono l'appartenenza "almeno o al più" ad una classe.

La regola decisionale descrive quindi gli oggetti secondo il principio di preferenza, formalmente:

$$\text{If } f(x, c_1) \geq r_{c1} \text{ and } \dots \text{ and } f(x, c_P) \geq r_{cP} \text{ then } x \geq D_1^{\geq} \text{ or } \dots \text{ or } x \geq D_k^{\geq}$$

Analogamente all'approccio classico gli elementi delle approssimazioni inferiori supportano regole certe, le approssimazioni superiori regole possibili, le frontiere regole incerte.

Le regole certe derivanti dalle unioni inferiori CI_l^{\geq} consentono di ordinare gli oggetti indicando il limite inferiore dell'ordinamento con relazioni di tipo:

$$\text{If } x_{c1} \geq r_{c1} \text{ and } \dots \text{ and } x_{cP} \geq r_{cP} \text{ then } x \in CI_l^{\geq}$$

Le regole certe derivanti dalle unioni inferiori CI_l^{\leq} consentono di ordinare gli oggetti indicando il limite superiore dell'ordinamento con relazioni di tipo:

If $x_{c1} \leq r_{c1}$ and ... and $x_{cP} \leq r_{cP}$ then $x \in Cl_t^{\leq}$

Le regole possibili derivanti dalle unioni superiori di Cl_t^{\geq} assegnano con ambiguità il limite inferiore:

If $x_{c1} \geq r_{c1}$ and ... and $x_{cP} \geq r_{cP}$ then x possibly belongs to Cl_t^{\geq}

Le regole possibili derivanti dalle unioni superiori di Cl_t^{\leq} assegnano con ambiguità il limite superiore:

If $x_{c1} \leq r_{c1}$ and ... and $x_{cP} \leq r_{cP}$ then x possibly belongs to Cl_t^{\leq}

Le regole approssimate derivanti dalle frontiere restituiscono regole incerte del tipo:

If $x_{c1} \leq r_{c1}$ and ... and $x_{cP} \leq r_{cP}$ then $x \in Cl_s \cup Cl_{s+1} \cup \dots \cup Cl_t$

La rappresentatività delle regole è indicata anche nell'approccio di dominanza dai valori *support*, *cover* e *strength*.

Un oggetto $x \in U$ supporta (*support*) una regola di decisione se la sua descrizione coincide sia con la parte condizionale che con quella decisionale.

Una regola di decisione copre (*cover*) un oggetto x se la descrizione di x coincide almeno con la parte condizionale della regola.

Una regola di decisione è migliore di un'altra se il valore di forza (*strength*), cioè il numero di oggetti che supportano la regola $\left(\frac{\text{Support}}{\text{Low - approximation}} \right)$ è maggiore.

Esempio 15

In accordo con i dati proposti e la procedura di definizione delle regole di assegnazione si ottengono le regole in Tabella B13. Dove il valore *support* indica il numero di oggetti che sostengono la regola, mentre il valore di *strength* indica il rapporto tra il numero di oggetti che definiscono la regola ed il numero di oggetti appartenenti al *P-lower* della classe di definizione. Ad esempio, la prima regola $q_3 \leq 1 \Rightarrow D$ at most 2 è supportata da tre oggetti (x_2, x_7, x_{10}) ed ha un valore di forza

relativa $3/3=100\%$. La regola $q_2 \geq 2 \Rightarrow D \text{ at least } 2$ è invece supportata da quattro oggetti (x_2, x_4, x_7, x_{10}) ed ha un valore di forza relativa $4/6=67\%$.

Tabella B13. Regole di decisione (*all rules on examples algorithm*).

N°	Condizione	Decisione	Support	Relative strength (%)
1	$q_3 \leq 1$	d at most 2	3	100%
2	$q_3 \geq 4$	d at least 3	2	100%
3	$q_1 \geq 3$	d at least 2	3	50%
4	$q_2 \geq 2$	d at least 2	4	66,67%
5	$q_1 \leq 2 \ \& \ q_3 \leq 3$	d = 3 or 2 or 1	5	100%
6	$q_3 \leq 3 \ \& \ q_3 \geq 2$	d = 3 or 2 or 1	5	100%
7	$q_2 \leq 1 \ \& \ q_3 \leq 3$	d = 3 or 2 or 1	4	100%

Appendice C: I modelli *Majority Rule Sorting* ed ELECTRE TRI

I Introduzione

Iniziamo questo appendice introducendo il così detto *Problem Statements*, ovvero la scelta del metodo cui ci si pone davanti ad un problema di decisione. Un problema di decisione (Colorni e Tsoukiàs, 2013) consiste nell'individuare un'appropriata partizione del set di alternative, secondo le necessità del decisore. Partizionare il set di alternative significa stabilire un insieme di classi di equivalenza alle quali associare gli elementi del set, le alternative. Le classi di equivalenza si possono distinguere per due ragioni: sono ordinate o no e sono pre-determinate o no (Figura C1).

- Per classi di equivalenza ordinate (generalmente *weak order*) si intendono classi che vantano un ordinamento transitivo e completo, cioè una classe può essere ritenuta migliore o uguale alle altre; un esempio sono le classi di merito. Quando le classi non sono ordinate ci riferiamo invece ad elementi su scala nominale, come appartenenza ad un colore o forma, una diagnosi medica o la diagnosi di un problema.
- Per classi pre-determinate ci riferiamo invece a classi derivanti da norme, standard, profili, ecc., come l'assegnazione di un voto da A ad F. Classi non pre-determinate sono invece *clustering* per attributi o classifiche, come il gruppo dei primi 100 o la classifica d'arrivo.

Figura C1. Basic Problem Statements (fonte: Colorni eTsoukiàs, 2013).

	Pre-determinate	Non Pre-determinate
Ordinate	Rating	Ranking
Non Ordinate	Assignment	Clustering

Diventa chiaro che il nostro problema si sviluppa sulla linea delle classi di equivalenza ordinate e pre-determinate, il nostro scopo è infatti quello di classificare le alternative appartenenti ad una tra le cinque/quattro classi di giudizio da Insufficiente a Ottimo; trattiamo quindi un problema di *Rating*. Brevemente, le procedure di ordinamento per classi si fondano sulla primitiva relazione di preferenza $a \geq b$ (a è ritenuta dal decisore almeno buona quanto b) e

quindi dalla relazione di preferenza operativa per cui $a \succ b$, significa che a è maggiore o uguale a b , ovvero che $a \in A$ e $b \in B : A \succ B$.

Le procedure di analisi multicriteriale (di *rating*) hanno lo scopo di assegnare le alternative alle categorie selezionate, lo fanno attraverso l'analisi delle primitive di ogni attributo e definiscono la classe di appartenenza secondo una regola di aggregazione. Nel metodo che andremo ad esaminare questa è una regola di maggioranza (ponderata) che somma l'importanza dei criteri per cui l'alternativa è almeno buona quanto la classe di appartenenza esaminata.

Tra i modelli di analisi di *rating* l'ELECTRE TRI (*ELimination Et Choix Traduisant la REalité*) è forse il più comune, lo andremo ad analizzare insieme ad una sua versione semplificata, l'MR-Sort (*Majority Rule Sorting model*).

II Concetti base dell'ELECTRE

La famiglia dei metodi ELECTRE (Yu, 1992; Roy e Bouyssou, 1993)¹⁰¹ nasce come trasposizione della procedura di voto di Condorcet nella *decision analysis* (Bouyssou *et al.*, 2000, p.129), sotto la famiglia dei metodi *outranking* (ELECTRE fu il primo metodo *outranking*), ovvero dei metodi costituiti da una procedura di confronto a coppie tra gli elementi e di associazione per relazioni binarie di concordanza e non discordanza. Come anticipato il principio del modello ELECTRE è che tra due alternative a e b , la prima è dichiarata “di grado superiore” (*outranks*) alla seconda, se ci sono abbastanza ragioni a supporto dell'affermazione e non ci sono sufficienti ragioni che sostengono il contrario. La logica del modello ELECTRE si fonda su una modellizzazione delle preferenze nella quale definiamo due relazioni binarie S_i e V_i per ogni dimensione:

$$a S_i b \Leftrightarrow g_i(a) \geq g_i(b)$$

$$a V_i b \Leftrightarrow g_i(a) > g_i(b) + \tau_i$$

dove τ_i indica un valore soglia considerato dal decisore “ampio”. La prima relazione esprime il fatto che a sia almeno buona come b per la dimensione (attributo) g_i . La seconda relazione indica che a è molto più buona di b per la

¹⁰¹ Yu, W.: *Aide multicritère la décision dans le cadre de la problématique du tri: méthodes et applications*. PhD thesis, LAMSADE, Université Paris Dauphine, Paris (1992).

Roy, B., Bouyssou, D.: *Aide multicritère la décision: méthodes et cas*. Economica Paris (1993).

dimensione g_i , poiché la differenza tra i valori eccede la soglia τ_i . La relazione binaria, chiamata *outranking relation*, è costruita sulla classe A. Se consideriamo che a è di grado superiore rispetto a b (a outranks b), se la coalizione di criteri per cui a è migliore di b è sufficientemente larga (concordanza) e se b non è molto migliore di a in una o più dimensioni (non discordanza). Per definire una “larga coalizione di criteri”, un coefficiente di importanza (peso) w_i è associato ad ogni criterio, la coalizione di criteri è quindi misurata dalla somma dei coefficienti di importanza per gli attributi per cui $a \succ b$, se il risultato è maggiore di un valore soglia λ , chiamato soglia di concordanza (*concordance threshold* o soglia di maggioranza). Formalmente, la relazione di *outranking* dipende dai valori di g , dai pesi w e dai valori soglia τ e dalla soglia di maggioranza λ . Formalmente:

$$a \succ (g, w, \tau, \lambda) b \Leftrightarrow \sum_{i: a \succ b} w_i \geq \lambda$$

$$\exists_i: b \succ a$$

In altre parole, l’alternativa a è preferita a b se la somma dei pesi degli attributi per i quali a è meglio di b supera la soglia di maggioranza λ e non esiste un attributo i per cui b è molto migliore di a .

II.1 ELECTRE TRI: componenti principali

La versione “TRI” del modello ELECTRE si fonda sui concetti appena esposti, ed operativamente è applicato come segue:

Consideriamo X un insieme di alternative valutate con n criteri $F=\{1,2,\dots,n\}$. Definiamo come a_j il valore dell’alternativa $a \in X$ per il criterio j .

Ogni criterio è valutato secondo l’appartenenza ad una delle categorie (classi di equivalenza) identificate come C_h , per le quali $C_h \geq C_{h+1} \geq \dots \geq C_{h+m}$.

Il profilo di ogni categoria C_h è delimitato dai valori soglia b_{h-1} e b_h , per i quali vale la regola $b_{h-1,j} \leq b_{h,j}$. Dove per $b_{h,j}$ è indicato il valore soglia dell’attributo j per il profilo h .

Esempio 1

Dato l’insieme $X=\{a_1, a_2, a_3\}$ composto dalle 3 alternative a_x e dati otto criteri $J =\{1,2,\dots, 8\}$, possiamo definire la matrice delle prestazioni in Tabella C1.

Tabella C1. Matrice delle 3 alternative a e 8 criteri j .

	j_1	j_2	j_3	j_4	j_5	j_6	j_7	j_8
a_1	4	2	3	1	2	3	4	3
a_2	1	2	1	3	4	3	2	4
a_3	1	4	3	1	1	2	1	3

Definiamo 4 categorie $C_1 > C_2 > C_3 > C_4$, ogni profilo è delimitato dai valori soglia b_{12j} , b_{23j} e b_{34j} . Per semplicità li definiremo uguali per ogni attributo j ed equivalenti a: $h_{12} \leq 2$, $h_{23} \leq 3$, $h_{34} \leq 4$. Ad esempio, il criterio j_1 dell'alternativa a_1 sarà classificato in C_4 , il criterio j_2 in C_2 , j_3 in C_3 , j_4 in C_1 e così via. Poiché i criteri di un'alternativa possono non essere classificati tutti nella stessa classe occorre definire una procedura di assegnazione per l'intera alternativa.

II.II Assegnazione delle alternative

Per assegnare un'alternativa a ad una classe di priorità C_h si applicano le seguenti due regole:

- Se un alternativa a ha uguale o maggiore priorità di b_{h-1} , deve essere classificato almeno nella classe di priorità C_h .
- se b_h ha uguale o maggiore priorità di a , l'alternativa a deve essere classificata al più nella classe di priorità C_h .

Formalmente:

$$a \in C_h \Leftrightarrow a P b_{h-1} \wedge b_h P a$$

dove P indica la relazione binaria di *outranking* "ha uguale o maggiore priorità di". La relazione P è basata sul principio di concordanza-discordanza, per il quale, un alternativa a è limitata al profilo b_h se c'è una significativa coalizione di criteri per cui " a ha la stessa priorità del profilo b_h " (principio di concordanza) e non ci sono significati opposizioni che sostengono il contrario (principio di discordanza).

Formalmente:

$$a P b_h \Leftrightarrow C(a, b_h) \wedge \neg D(a, b_h)$$

dove:

Dario Canu

Camminabilità, perché e come misurarla: un modello e uno studio empirico di validazione della misura
Dottorato in Architettura e Ambiente, XXIX Ciclo, Università degli studi di Sassari

- $C(a, b_h)$ indica che c'è una maggioranza di criterio che supportano la proposizione che a “ha priorità almeno quanto” (outranks) b_h .
- $D(a, b_h)$ indica che c'è una forte opposizione, definita veto, per cui a “ha priorità almeno quanto” b_h .

In accordo con Roy (1968)¹⁰², $C(x,y)$ e $D(x,y)$ sono formalmente:

$$C(x,y) \Leftrightarrow \frac{\sum_{i \in H(x,y)} w_i}{\sum_{j=1}^n w_j} \geq \lambda$$

$$D(x,y) \Leftrightarrow \exists h_i : h_i(y) - h_i(x) > v_i$$

dove:

- $h_i, i = 1, \dots, n$ sono i criteri;
- w_i sono i coefficienti di importanza (pesi) associati ad ogni criterio;
- $h_i(x)$ è la valutazione di x per il criterio h_i ;
- $H(x,y)$ è il set di criteri per cui x ha lo stesso o maggior valore di y , $h_i(x) \geq h_i(y)$;
- λ è la soglia di maggioranza;
- v_i è la soglia di veto del criterio h_i .

Questo significa che un'alternativa è assegnata ad una categoria C_h se le sue caratteristiche sono almeno buone quando la soglia inferiore della categoria b_h e peggiori della soglia superiore b_{h+1} (che delimita la classe C_{h+1}) per una maggioranza pesata di criteri. Formalmente:

$$a \in C_h \equiv a S b_h \Leftrightarrow \sum_{j: a_j \geq b_{h,j}} w_j \geq \lambda$$

dove w_j per $j \in F$ sono i pesi (non negativi) associati ad ogni criterio che soddisfano la condizione di normalizzazione $\sum_{j \in F} w_j = 1$.

Perciò, le condizioni che definiscono un'alternativa $a \in X$ assegnata alla categoria C_h sono:

¹⁰² Roy B.: Classement et choix en présence de points de vue multiples (la méthode ELECTRE), *La Revue d'Informatique et de Recherche Opérationnelle* (RIRO) (8), (1968) 57–75.

$$\sum_{j: a_j \geq b_{h-1,j}} w_j \geq \lambda$$

$$\sum_{j: a_j \geq b_{h,j}} w_j < \lambda$$

Esempio 2

Dati i valori dell'esempio 1, per poter procedere è necessario definire i valori dei pesi w_j di ogni criterio e la soglia di maggioranza λ .

Supponendo di avere i pesi per ogni criterio w_j i seguenti valori: 0, 0,2, 0,1, 0,15, 0,15, 0,1, 0,15, 0,15 e di aver definito la soglia di maggioranza $\lambda=0,6$, possiamo attribuire l'appartenenza di ogni alternativa ad una classe.

a_1 è classificata in C_2 perché la somma dei pesi dei criteri per cui a_1 appartiene a C_2 è maggiore della soglia di maggioranza ($w_1+w_2+w_3+w_5+w_6+w_7+w_8 \geq \lambda$), e la somma dei pesi dei criteri per cui a_1 appartiene a C_3 è inferiore alla soglia di maggioranza ($w_1+w_3+w_6+w_7+w_8 < \lambda$).

Allo stesso modo a_2 appartiene alla classe C_2 . Infatti tutti i criteri includono a_2 in C_1 , una maggioranza pesata dello $0,9 > \lambda$ sostiene che $a_2 \in C_2$, mentre $a_2 \in C_3$ per un valore di $0,55 (< \lambda)$ e $a_2 \in C_4$ solo per $0,3$. a_3 è invece classificato in C_1 , le somme dei pesi di ogni coalizione per le quattro classi sono: 1,0, 0,55, 0,45 e 0,2.

II.III Precisazioni sul modello MR-Sort

Abbiamo detto che la procedura *Majority Rule Sorting model* è figlia dell'ELECTRE TRI poiché nasce come una sua versione semplificata¹⁰³. Difatti applicare questa procedura significa valutare un set di alternative con la procedura dell'ELECTRE TRI che abbiamo visto sino ad ora, ignorando i passaggi che seguono. Significa escludere dal processo di valutazione il test di non discordanza per cui un'alternativa non può essere classificata in una certa classe se ci sono alcune regole in contrasto con la classificazione iniziale.

¹⁰³ Sobrie O., Mousseau V. Pirlot M. (2013) *Learning a Majority Rule Model from Large Sets of Assignment Examples*, in P. Perny, M. Pirlot, and A. Tsoukiás (Eds.): ADT 2013, LNAI 8176, pp. 336–350.

II.IV Regole di veto

Il principio di discordanza del modello ELECTRE è tradotto in maniera formale dall'applicazione delle regole di veto, che sono delle condizioni che impediscono alcune classificazioni al verificarsi di determinate condizioni. Nel processo di classificazione, se un'alternativa supera il test di concordanza, cioè la maggioranza pesata degli indicatori supera la soglia λ , l'alternativa è confrontata con le regole di veto, se il test di discordanza è superato l'alternativa appartiene alla classe definita, altrimenti subisce un processo di revisione in accordo con le regole di veto. Formalmente le regole di veto possono essere descritte secondo regole del tipo *if, then*:

$$g_h > J_h \Rightarrow a \in C_h$$

dove g_h indica il valore g del criterio h e J_h indica il valore soglia per il criterio. Significa che $a \in C_h$ se il criterio h ha un valore maggiore del livello massimo consentito J .

Esempio 3

Ipotizziamo di integrare le seguenti regole di veto per ricalcolare l'appartenenza delle quattro alternative ad una delle quattro classi C_h . Introduciamo tre regole di veto:

$$\begin{aligned} g_1 &\geq 4 \Rightarrow a \in C_4 \\ g_2 &\geq 4 \Rightarrow a \in C_3 \\ g_3 &< 2 \Rightarrow a \in C_1 \vee C_2 \\ g_5 \wedge g_6 &\geq 3 \Rightarrow a \in C_3 \end{aligned}$$

Secondo la classificazione precedente $a_1 \in C_2$, la prima regola di veto impone però che a_1 sia classificata in C_4 .

L'alternativa C_2 risulta invece correttamente classificata, infatti la regola di veto per cui se g_3 è minore di 2 la classe appartiene al più a C_2 è rispettata.

Infine, a_3 , secondo la quarta regola di veto, deve essere classificata in C_3 a differenza di quanto stabilito dal test di concordanza che classificava l'alternativa in C_2 .

Appendice D: Le reti neurali

Le reti neurali artificiali (*artificial neural networks*) sono composte da procedure statistiche di apprendimento di un modello e sono ispirate dalle reti neurali biologiche. Sono usate nel *machine learning* per il *pattern recognition*, l'analisi dei dati, il controllo o per la classificazione e assegnazione.

Una rete neurale è composta da due principali elementi, i neuroni e le connessioni inter-neurali. Nella rete biologica i neuroni hanno la funzione di produrre impulsi elettrici contenenti informazioni, le connessioni inter-neurali quella di collegare più neuroni per il passaggio delle informazioni. Ogni sinapsi ha un "peso" che determina l'intensità dell'impulso proveniente dal neurone precedente, se la forza è sufficiente il successivo neurone viene attivato e trasmette le informazioni ad altri neuroni. Ogni neurone è quindi stimolato da un impulso di *input* e, se l'intensità è sufficiente alla sua attivazione, restituisce un impulso di *output*. Analogamente, in una rete artificiale il funzionamento è modellizzato da una struttura di neuroni come funzioni lineari o, generalmente, non lineari (sigmoidea logistica, tangente iperbolica, soglia) che trasformano gli *input* ($x_1, x_2, \dots, x_i, \dots, x_n$) in output y , trasmessi dalle connessioni inter-neurali agli altri neuroni lungo la rete. L'attivazione di ogni connessione successiva è funzione dell'importanza della connessione e dell'intensità del segnale definito "peso sinattico". Il rapporto tra le interazioni di ogni percorso e dei loro pesi darà luogo ad un differente risultato di *output*.

Proprio come una rete neurale biologica quella artificiale definisce il suo funzionamento istruendosi dall'ambiente; così come si impara a distinguere – per rimanere in campo urbanistico – una strada carrabile da una pedonale, attraverso la loro osservazione e l'individuazione degli elementi che caratterizzano l'una o l'altra, una rete neurale artificiale, a partire da un set di dati di insegnamento, costruisce una rete che ne descrive il funzionamento ed che è capace di prevedere i risultati di nuovi dati. Esistono diversi tipi di reti neurali artificiali, per diversi scopi, ma la divisione fondamentale sta nel tipo di apprendimento, supervisionato o non-supervisionato. Nel primo caso si addestra la rete con degli esempi che contengono l'*input* e l'*output* associato, mentre nel secondo caso la rete deve essere in grado di estrarre delle informazioni di similitudine tra i dati forniti in *input* (senza associazioni con *output* desiderati) al fine di classificarli in categorie. In base al tipo di utilizzo si può distinguere tra memorie associative (es.

associazioni tra pattern di immagini), simulatori di funzioni matematiche o classificatori. Per questo lavoro ci interessa analizzare l'apprendimento supervisionato per classificatori, brevemente descritto di seguito.

I Struttura di una rete neurale

Una rete neurale artificiale è generalmente costituita da tre gruppi principali, un gruppo di neuroni di *input*, uno di *output* e uno o più gruppi definiti "nascosti" di neuroni intermedi. Lo strato principale, di *input*, contiene i predittori, cioè le variabili che descrivono le alternative che si stanno valutando, lo strato di *output* contiene invece le risposte, cioè la categoria di appartenenza delle alternative. Lo strato intermedio contiene unità o nodi non osservabili, il valore di ognuna delle unità nascoste dipende da una funzione dei predittori.

Esempio 1

Ad esempio, una rete neurale composta da due neuroni di *input* i_1 e i_2 e due di *output* o_1 e o_2 , con un livello nascosto composto da due neuroni h_1 e h_2 è rappresentata in Figura D1. Ogni w_i indica il peso sinattico della connessione corrispondente.

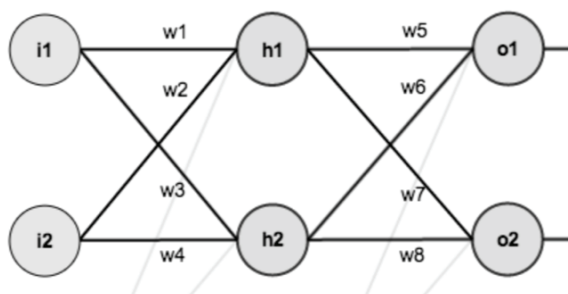


Fig. D1 Esempio di rete neurale.

I.I Definizione della struttura

Una rete neurale definisce la sua struttura a partire da un set di dati (di *training*) che istruisce la rete attraverso un processo di ottimizzazione dei valori che compongono il sistema. Istruire una rete significa calibrare i pesi sinattici ripetendo più volte i due passi chiave, propagazione in avanti e propagazione indietro (*forward propagation* e *back propagation*).

Nella *forward propagation* a partire da una situazione casuale si applicano una serie di pesi per i dati in ingresso e si calcolano i valori di uscita. Nella *back propagation* si misurano gli errori marginali degli *output* e si modificano i pesi secondo un processo di minimizzazione dell'errore. Generalmente il processo è ripetuto un certo numero di volte o sino a che la variazione dell'errore non scende sotto una certa soglia.

Esempio 2

Data la rete in Figura D1 possiamo assegnare in maniera casuale i valori di *input*, *output* e dei pesi, includendo i valori di *bias* b_1 e b_2 per lo strato intermedio e di *output*, come rappresentato in Figura D2, che riporta lo schema e la tabella dei valori.

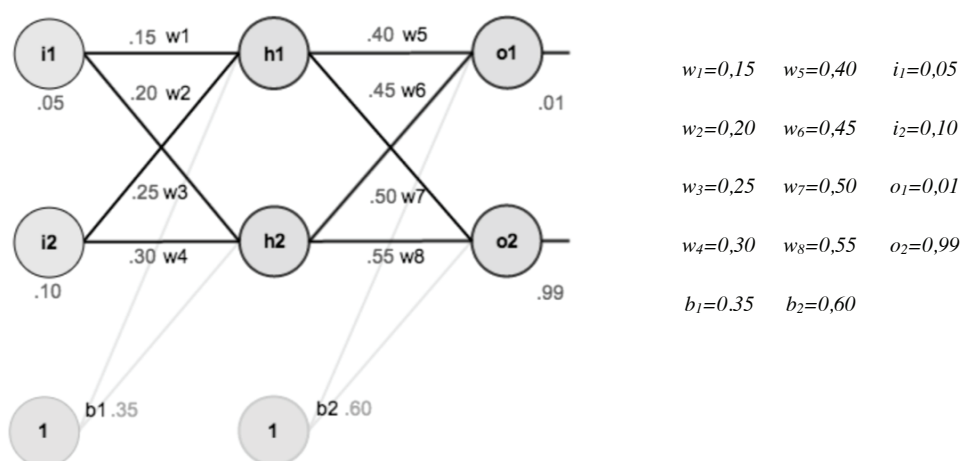


Fig. D2 Primo *step* della costruzione di una rete neurale, assegnazione casuale dei valori.

II *forward propagation*

Il primo passo consiste nella stima dei valori di *output* generati dalla rete così come strutturata. Calcoliamo quindi il valore di ingresso *net* per ogni neurone del livello nascosto, dalla somma dei prodotti dei valori di ogni neurone *input* per il peso della connessione, compresi i *bias*.

Il valore di *output* *out* è influenzato dall'attivazione del neurone ed è calcolato applicando – ad esempio – la funzione logistica.

Lo stesso processo è ottenuto per i neuroni risposta, sapendo che il valore *net* per i neuroni di *output* è ottenuto dalla somma dei prodotti dei valori di out_h moltiplicati per i pesi sinattici delle connessioni da *h* ad *o*.

Esempio 3

Secondo i dati della rete neurale in esempio, possiamo calcolare il valore in ingresso per il neurone h_1 secondo la formula:

$$net_{h1} = w_1 \times i_1 + w_2 \times i_2 + b_1 \times 1 = 0,15 \times 0,05 + 0,2 \times 0,1 + 0,35 \times 1 = 0,3775$$

applicando la funzione di attivazione si ottiene il valore di output:

$$out_{h1} = 1/(1+e^{-net_{h1}}) = 1/(1+e^{-0,3775}) = 0,5932$$

perciò, secondo lo stesso processo:

$$net_{o1} = w_5 \times out_{h1} + w_6 \times out_{h2} + b_2 \times 1 = 1,1059$$
$$out_{o1} = 1/(1+e^{-net_{o1}}) = 0,7514$$

II.I Calcolo dell'errore totale

L'errore di previsione del modello è calcolato secondo la formula dell'errore quadratico medio (*squared error function*) per cui:

$$E_{tot} = \sum 1/2 (target - output)^2.$$

Esempio 4

in questo caso l'errore totale è calcolato dalla somma degli errori E_{o1} e E_{o2} rispettivamente per i due neuroni *output*. Ognuno dei quali è calcolato secondo la formula:

$$E_{o1} = \sum 1/2 (target_{o1} - output_{o1})^2 = 1/2(0,01-0,7514)^2 = 0,2748$$

Secondo lo stesso procedimento $E_{o2} = 0,0236$, tale che:

$$E_{tot} = E_{o1} + E_{o2} = 0,2748 + 0,0236 = 0,2984$$

III Back propagation (output layer)

A questo punto ciò che serve è aggiornare i pesi neurali e i pesi sinattici affinché lo scarto tra il valore di *output* del modello e il valore reale osservato dagli esempi

di assegnazione sia il più simile possibile, ovvero che l'errore sia minore, riducendo al minimo l'errore per ogni neurone di *output* e dell'intera rete.

Serve quindi valutare l'influenza di ogni interconnessione neurale sull'errore totale, ovvero $\partial E_{tot}/\partial E_{w_x}$ che corrisponde alla derivata parziale di E_{tot} rispetto a w_x .

Per risolvere l'equazione, applicando la regola della catena, si ottiene che:

$$\frac{\partial E_{tot}}{\partial w_x} = \frac{\partial E_{tot}}{\partial out_o} \times \frac{\partial out_o}{\partial net_o} \times \frac{\partial net_o}{\partial w_x} = - (target_{o1} - out_{o1}) \times out_{o1} (1 - out_{o1}) \times out_{h1}$$

Esempio 5

Possiamo ad esempio calcolare l'influenza di w_5 secondo i dati dell'esempio. Per risolvere l'equazione è necessario esplicitare ogni componente:

$$\partial E_{tot}/\partial out_{o1} = 2 \times 1/2 (target_{o1} - out_{o1}) 2-1 \times -1 + 0 = 0,7414$$

$$\partial out_{o1}/\partial net_{o1} = out_{o1} (1 - out_{o1}) = 0,1868$$

$$\partial net_{o1}/\partial w_5 = 1 \times out_{h1} \times w_5(1-1) + 0 + 0 = out_{h1} = 0,5933$$

Per cui:

$$\partial E_{tot}/\partial E_{w_5} = 0,7414 \times 0,1868 \times 0,5933 = 0,0822$$

III.I Riduzione dell'errore

Per ridurre l'errore si sottrae dal peso w_x il valore ottenuto dal calcolo precedente (opzionalmente) moltiplicato per un tasso di apprendimento *learning rate*, eta.

Esempio 6

In questo caso, applicando un tasso $\eta=0,5$ il valore corretto di w_5 è calcolato secondo la formula:

$$w_5^+ = w_5 - \eta \times \partial E_{tot}/\partial w_5 = 0,4 - 0,5 \times 0,0822 = 0,3589$$

IV Back propagation (hidden layer)

La procedura precedente, di riduzione dell'errore del valore dei pesi, è applicata solo dopo aver calcolato l'errore anche per i livelli nascosti. Il funzionamento è

simile se non per alcune differenze causate dall'influenza multipla che i neuroni (nascosti) hanno sui neuroni di *output*.

Esempio 7

Infatti, per il caso esempio, il valore di out_{h1} influenza sia out_{o1} sia out_{o2} , perciò:

$$\partial E_{tot}/\partial out_{h1} = \partial E_{o1}/\partial out_{h1} + \partial E_{o2}/\partial out_{h1}$$

dove:

$$\partial E_{o1}/\partial net_{o1} = \partial E_{o1}/\partial out_{o1} \times \partial out_{o1}/\partial net_{o1} = 0,7414 \times 0,1868 = 0,1385$$

$$\partial net_{o1}/\partial out_{h1} = w_5 = 0,40$$

$$\partial E_{o1}/\partial out_{h1} = \partial E_{o1}/\partial net_{o1} \times \partial net_{o1}/\partial out_{h1} = 0,1385 \times 0,4 = 0,0554$$

secondo la stessa procedura,

$$\partial E_{o2}/\partial out_{h1} = -0,0190, \text{ per cui } \partial E_{tot}/\partial out_{h1} = 0,0363.$$

IV.I Calcolo del valore di influenza

A questo punto possiamo calcolare il valore di influenza $\partial E_{tot}/\partial E_{wx}$ di ogni connessione tra il livello nascosto e quello di *input* (che corrisponde alla derivata parziale di E_{tot} rispetto a w_x) secondo la formula.

$$\frac{\partial E_{tot}}{\partial w_x} = \frac{\partial E_{tot}}{\partial out_h} \times \frac{\partial out_h}{\partial net_h} \times \frac{\partial net_h}{\partial w_x}$$

Esempio 8

Ottenuto $\partial E_{o1}/\partial out_{h1}$ ci serve conoscere $\partial out_{h1}/\partial net_{h1}$ e $\partial net_{h1}/\partial w$ per ogni peso.

$$\partial out_{h1}/\partial net_{h1} = out_{h1}(1-out_{h1}) = 0,5932(1-0,5932) = 0,2413$$

$$\partial net_{h1}/\partial w_1 = i_1 = 0,05$$

Mettendo insieme i valori ottenuti secondo la formula di $\partial E_{tot}/\partial w_1$ otteniamo $0,0363 \times 0,2413 \times 0,05 = 0,0004$.

IV.II Riduzione dell'errore

Per ridurre l'errore, come nel caso precedente, si sottrae dal peso w_x il valore di $\partial E_{tot}/\partial w_x$ per eta.

Esempio 9

Applicando un tasso $\eta=0,5$ il valore corretto di w_1 è calcolato secondo la formula:

$$w_1^+ = w_1 - \eta \times \partial E_{tot}/\partial w_1 = 0,15 - 0,5 \times 0,0004 = 0,1497$$

V Ripetizione del processo

Completata la fase di misura dell'errore per ogni peso possiamo aggiornarli. Rispetto ai valori iniziali di *input* $i_1=0,05$ e $i_2=0,10$ l'errore originale della rete era 0,29837. Dopo il primo turno di *back propagation* l'errore totale si è ridotto a 0,29103. Dopo 10.000 volte, l'errore si sarà ridotto a 0,00003. I valori di *output* passano quindi da una stima di $out_{o1}=0,7514$ e $out_{o2}=0,7729$ ottenuta dai valori casuali a 0,0159 e 0,9841 rispetto a dei valori reali di 0,01 e 0,99.

Bibliografia

- Apparicio, P., Cloutier, M.S., Shearmur, R. (2007). The case of Montreal's missing food deserts: Evaluation of accessibility to food supermarkets. *International Journal of Health Geographies*, 6(4).
- Bejleri, I., Steiner, R. L., Fischman, A., Schmucker, J. M. (2010). Using GIS to analyze the role of barriers and facilitators to walking in children's travel to school. *URBAN DESIGN International*, 16(1): 51–62.
- Benenson, I., Martens, K., Rofé, Y., Kwartler, A. (2011). Public transport versus private car GIS-based estimation of accessibility applied to the Tel Aviv metropolitan area. *Annals of Regional Science*, 47(3): 499-515.
- Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*, Clarendon Press, Oxford.
- Bertolini, L. (1999) Spatial development patterns and public transport: the application of an analytical model in the Netherlands, *Planning Practice and Research*, 14(2): 199-210.
- Bertolini, L., Djist, M. (2003) Mobility environments and network cities, *Journal of Urban Design*, 8(1): 27-43.
- Blečić, I., Cecchini, A., Congiu, T., Fancello, G., Trunfio, G.A. (2014a) Walkability Explorer. An evaluation and design support tool for walkability, *Computational Science and Its Applications*, 8582: 511-521.
- Blečić, I., Cecchini, A., Congiu, T., Fancello, G., Trunfio, G.A. (2015a) Evaluating walkability: a capability-wise planning and design support system, *International Journal of Geographical Information Science*, 8816: 1-25.
- Blečić, I., Cecchini, A., Congiu, T., Pazzola, M., Trunfio, G.A. (2013) A Design and Planning Support System for Walkability and Pedestrian Accessibility, *Computational Science and Its Applications*, 7974: 284-293.
- Blečić, I., Cecchini, A., Fancello, G., Talu, V., Trunfio, G. A. (2015b). Walkability and urban capabilities: evaluation and planning decision support, *Territorio Italia*, 1: 51-66.
- Blečić, I., Cecchini, A., Minchilli, M., Talu, V. (2014b). Progettare la città di prossimità per promuovere le « capacità urbane » degli abitanti svantaggiati. *International Conference Virtual City and Territory*, 1240-1248.
- Bouyssou, D., Marchant, T. (2007a) An axiomatic approach to noncompensatory sorting methods in MCDM, I: The case of two categories. *European Journal of Operational Research*, 178(1): 217–245.
- Bouyssou, D., Marchant, T. (2007b) An axiomatic approach to noncompensatory sorting methods in MCDM, II: More than two categories. *European Journal of Operational Research*, 178(1): 246–276.

- Bouyssou D. Marchant T. Pirlot M. Perny P. Tsoukiàs A. Vincke P. (2000), *Evaluation and Decision Models - A critical perspective*, Kluwer Academic Publishers, Boston/London/Dordrech.
- Brüggemann, R., Patil, G.P. (2011) *Ranking and Prioritization for Multi-indicator Systems*, Springer, New York.
- Buck, C., Pohlabeln, H., Huybrechts, I., De Bourdeaudhuij, I., Pitsiladis, Y., Reisch, L., Pigeot, I. (2011) Development and application of a moveability index to quantify possibilities for physical activity in the built environment of children, *Health and Place*, 17(6): 1191–1201.
- Burns, L.D. (1979) *Transportation, Temporal, and Spatial Components of Accessibility*. Lexington Books, Lexington, MA.
- Carr, L. J., Dunsiger, S. I., Marcus, B. H. (2010) Walk Score as a global estimate of neighborhood walkability. *American Journal of Preventive Medicine*, 39(5): 460–463.
- Cecchini, A. Talu, V. (2011) Camminare (e pedalare) per trasformare la città, *TeMA*, 4(4): 99-108.
- Cecchini, A., Musci, E. (a cura di) (2008) *Differenti? è indifferente. Capire l'importanza delle differenze culturali e fare in modo che non ci importi*, Edizioni la meridiana, Molfetta.
- Cerin, E., Macfarlane, D.J., Ko, H.H., Chan, K.C.A. (2007) Measuring perceived neighbourhood walkability in Hong Kong, *Cities*, 24(3): 209–217.
- Chakhar, S., Saad, I. (2012a) Dominance-based rough set approach for groups in multicriteria classification problems. *Decision Support Systems*, 54(1): 372–380.
- Chang, H.S., Liao, C.H. (2011) Exploring an integrated method for measuring the relative spatial equity in public facilities in the context of urban parks, *Cities*, 28: 361-371.
- Chin, H.C., Foong, K.W. (2006) Influence of school accessibility on housing values, *Journal of Urban Planning and Development*, 132(3): 120-129.
- Colomi, A., Tsoukiàs, A. (2013) What is a decision problem? Preliminary statements, *Lecture Notes in Computer Science*, 8176: 139–153.
- Comber, A., Brunson, C., Green, E. (2008) Using a GIS-based network analysis to determine urban green space accessibility for different ethnic and religious groups, *Landscape and Urban Planning*, 86(1): 103-114.
- Coutts, C. (2008) Greenway accessibility and physical-activity behavior, *Environment and Planning B*, 35: 552-563.
- Cowan, R. (2005). *The Dictionary of Urbanism*. Streetwise Press, Wiltshire.
- Crompton, J.L., Wicks, B.E. (1988) Implementing a preferred equity model for the delivery of leisure services in the US context, *Leisure Studies*, 7: 287-304.

- Cummins, S., Macintyre, S. (2002) A systematic study of an urban foodscape: the price and availability of food in Greater Glasgow, *Urban Studies*, 39: 2115-2130.
- Dalvi, M.Q. (1978) Behavioural modelling accessibility, mobility and need: concepts and measurement. In: Hensher, D.A., Stopher, P.R., (a cura di), *Behavioural Travel Modelling*, London.
- Dalvi, M.Q., Martin, K. (1976) Estimate of non work trip demand: a disaggregated approach. In: Bonsall P., Dalvi, M.Q. and Hills P.J., (a cura di), *Urban Transportation Planning: Current Problems and Future Prospects*. Tunbridge Wells, Abacus Press.
- Diez Roux, A., Evenson, K., McGinn, A., Brown, D.G., Moore, L., Brines, S., Jacobs, D.R. (2007) Availability of recreational resources and physical activity in adults, *American Journal of Public Health*, 97: 493-499.
- Evers, C., Boles, S., Johnson-Shelton, D., Schlossberg, M. (2014) Parent safety perceptions of child walking routes, *Journal of Transport & Health*, 1: 108–115.
- Ewing, R., & Handy, S. (2009) Measuring the unmeasurable: Urban Design Qualities Related to Walkability, *Journal of Urban Design*, 14 (1): 65-84.
- Ewing, R., Handy, S., Crownson, R.C., Clemente, O., Winston, E. (2006) Identifying and Measuring Urban Design Qualities Related to Walkability, *Journal of Physical Activity and Health*, 3(1): 223-240.
- Fainstein, S. (2010) *The just city*, Cornell University, New York.
- Farrington, J., Farrington, C. (2005) Rural accessibility, social inclusion and social justice: towards conceptualisation, *Journal Transport Geography*, 13(1): 1-12.
- Filipucci, M. (2015) Primitive urbane. Analisi interpretativa dei processi figurativi dell'immagine della città, 159-168, In Aa. Vv. *Disegno & Città / Drawing & City: Cultura, Arte, Scienza, Informazione*, Gangemi Editore, Roma.
- Fleiss, J.L. (1981) *Statistical methods for rates and proportions* (2nd ed.), John Wiley, New York.
- Frau, G. (2015) Natural perceptual wayfinding for urban accessibility of the elderly with early-stage, *Journal of Technology for Architecture and Environment*, 1: 106-113.
- Furlan, R., Martone, D. (2001) *La conjoint analysis per la ricerca sociale e di marketing*, Franco Angeli, Milano.
- Gallimore, J.M., Brown, B.B., Wener, C.M. (2011) Walking routes to school in new urban and suburban neighborhoods: An environmental walkability analysis of blocks and routes. *Journal of Environmental Psychology*, 31: 184-191.
- Garb, Y., Levine, J. (2002) Congestion Pricing's Conditional Promise: Promotion of Accessibility or Mobility?, *Transport Policy*, 9(3): 179-188.

- Garin, N., Olaya, B., Miret, M., Ayuso-Mateos, J. L., Power, M., Bucciarelli, P., Haro, J. M. (2014). Built environment and elderly population health: a comprehensive literature review, *Clinical practice and epidemiology in mental health: CP & EMH*, 10: 103–115.
- Gauvin, L., Richard, L., Craig, C.L., Spivock, M., Riva, M., Forster, M., Laforest, S., Laberge, S., Fournel, M.C., Gagnon, H., Gagné, S., Potvin, L. (2005) From Walkability to Active Living Potential, an “Ecometric” Validation Study, *American Journal of Preventive Medicine*, 28(2): 126-133.
- Geurs, K.T., van Eck, R. (2001) *Accessibility measures: review and applications. Evaluation of accessibility impacts of land use transportation scenarios and related social and economical impact*. RIVM Report, Utrecht University.
- Geurs, K.T., van Wee, B. (2004). Accessibility of land-use and transport strategies: review and research directions, *Journal of Transport Geography*, 12: 127-140.
- Giergiczny, M., Kronenberg, J. (2014). From valuation to governance: Using choice experiment to value street trees, *Ambio*, 43(4): 492–501.
- Goldberger, A.S (1972) Structural equation Methods in the social sciences, *Econometrica*, 40, 979-1001.
- Greco, S., Matarazzo, B., Slowinshi, R. (1997) Rough set approach to multiple attribute choice and ranking problems’, *multiple criteria decision making*, 448: 318-329.
- Greco, S., Matarazzo, B., Slowinshi, R. (1999) Rough set approximation of a preference relation by dominance relations, *European journal of operational research*, 117(1): 63-83.
- Greco, S., Matarazzo, B., Slowinshi, R. (2001) Rough set theory for multicriteria decision analysis, *European journal of operational research*, 129(1): 1-47.
- Guo, Z., Loo, B.P.Y. (2013) Pedestrian environment and route choice: evidence from New York City and Hong Kong, *Journal of Transport Geography*, 28: 124-136.
- Gutiérrez, J. (2001) Location, economic potential and daily accessibility: an analysis of the accessibility impact of the high-speed line Madrid–Barcelona–French border, *Journal of Transport Geography*, 9: 229-242.
- Gutiérrez, J., Urbano, P. (1996) Accessibility in the European Union: the impact of the trans-European road network. *Journal of Transport Geography*, 4(1): 15-25.
- Hanák, T., Marović, I., Aigel, P. (2015) Perception of Residential Environment in Cities: a Comparative Study. *Procedia Engineering*, 117: 495 – 501.
- Handy, S.L., Niemeier, D.A. (1997). Measuring accessibility: an exploration of issues and alternatives, *Environment and Planning A*, 29: 1175-1194.
- Hanley, N., Mourato, S., Wright, R. E. (2001) Choice Modelling Approaches: a Superior Alternative for Environmental Valuation?, *Journal of Economic Surveys*, 15(3): 435–462.

- Hansen, W.G. (1959) How accessibility shapes land use, *Journal of American Institute of Planners*, 25(1): 73-76.
- Harmansah, Ö. (2007) The Archaeology of Mesopotamia: Ceremonial centers, urbanization and state formation in Southern Mesopotamia, In *Archaeologia: the world's most significant site and cultural treasures*, Frances Lincoln Publishers, London , 214-223.
- Harnik, P. (2003) *The Excellence City Park System: What Makes it Great and How to Get There*. The Trust for Public Land Pub, Washington, DC.
 Testo disponibile al sito: http://cloud.tpl.org/pubs/ccpe_excellentcityparks_2006.pdf (ultimo accesso 03/09/2015).
- Harvey, D. (1973) *Social justice and the city*, Edward Arnold Publishers Ltd, London.
 Trad. Ita. (1978) *Giustizia sociale e città, tesi liberarli, tesi socialiste*, Feltrinelli, Milano.
- Harvey, D. (2012) *Il capitalismo contro il diritto alla città*, Ombre Corte, Verona.
- Harvey, D. (2012). *Rebel cities. from the right to the city to the right to the urban revolution*. VersoBook, London, New York.
 Trad. Ita. (2013) *Città ribelli, i movimenti urbani dalla comune di Parigi a Occupy Wall Street*. ilSaggiatore, Milano.
- Hay, A.M. (1995) Concepts of equity, fairness and justice in geographical studies, *Transactions of the Institute of British Geographers*, 20: 500-508.
- Haykin, S. (1998) *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing, New York.
- Hensher, D.A. (2004) Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings, *Journal of Transport Economics and Policy*, 38(3): 425-446.
- Hillsdon, M., Panter, J., Foster, C., Jones, A. (2006) The relationship between access and quality of urban green space with population physical activity, *Public Health*, 120: 1127-1132.
- Iacono, M., Krizek, K.J., El-Geneidy, A. (2010). Measuring non-motorized accessibility: issues, alternatives, and execution, *Journal of Transport Geography*, 18: 133-140.
- Indovina, F. (2009) *Dalla città diffusa all'arcipelago metropolitano*, Franco Angeli, Milano.
- Ischia G. (2012) *La città giusta: idee di piano e atteggiamenti etici*, Donzelli Editore, Roma.
- Jacquet-Lagreze, E., Siskos, J. (1982) Assessing a set of additive utility functions for multicriteria decision-making, the UTA method, *European Journal of Operational Research*, 10(2): 151-164.

- Jacquet-Lagrèze, E., Siskos, Y. (2001). Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research*, 130(2): 233–245.
- Jensen, S. (2007) Pedestrian and Bicyclist Level of Service on Roadway Segments. *Transportation Research Record: Journal of the Transportation Research Board*, 2031: 43-51.
- Johnston, R.J., Gregory, D., Pratt, G., Watts M. (2000) *The Dictionary of Human Geography*, Blackwell Publishing, Oxford.
- Kadali, B.R., Vedagiri, P. (2015) Evaluation of pedestrian crosswalk level of service (LOS) in perspective of type of land-use, *Transportation Research Part A*, 73, 113–124.
- Kang, L., Xiong, Y., Mannering, F.L. (2013) Statistical analysis of pedestrian perceptions of sidewalk level of service in the presence of bicycles, *Transportation Research Part A*, 53: 10–21.
- Keeney, R.L. (1992) *Value-focused thinking: a path to creative decisionmaking*, Harvard University Press, Cambridge.
- Keeney, R.L. (1996), Value-focused thinking: Identifying decision opportunities and creating alternatives, *European Journal of Operational Research*, 92(3): 537–549.
- Khisty, C.J., (1994) Evaluation of pedestrian facilities: Beyond the level-of-service concept. *Transport Research Record* 1438, 45–50.
- Kinman, E.L. (1999) Evaluating health services equity at a primary care clinic in Chilimarca. *Bolivia Social Science & Medicine*, 49(5): 663-678.
- Knox, P.L. (1978) The intraurban ecology of primary medical care: patterns of accessibility and their policy implications, *Environment and Planning A*, 10: 415-435.
- Koenig, J.G. (1980) Indicators of urban accessibility: theory and application, *Transport*, 9: 145-172.
- Koh, P.P., Wong, Y.D. (2013) Influence of infrastructural compatibility factors on walking and cycling route choices, *Journal of Environmental Psychology*, 36: 202-213.
- Kunzmann, K.R. (1998) Planning for spatial equity in Europe, *International Planning Studies*, 3(1): 101-121.
- Lamíquiz, P.J., López-Domínguez, J. (2015) Effects of built environment on walking at the neighbourhood scale. A new role for street networks by modelling their configurational accessibility?, *Transportation Research Part A*, 74: 148–163.
- Landis, B.W., Vattikuti, V.R., Ottenberg, R.M, McLeod, D.S., Guttenplan, M. (2001) Modeling the roadside walking environment: a pedestrian level of service, *Transportation Research Record*, 1773: 82-88.
- Landis, J. R., Koch, G. G. (1977) The measurement of observer agreement for categorical data, *Biometrics*, 33(1): 159–174.

- Lau, J.C.Y., Chiu, C.C.H. (2003) Accessibility of low-income workers in Hong Kong. *Cities*, 20(3): 197-204.
- Lau, J.C.Y., Chiu, C.C.H. (2004) Accessibility of workers in a compact city: the case of Hong Kong, *Habitat International*, 28: 89-102.
- Lefebvre H. (2014) *Il diritto alla città*, Ombre Corte, Verona.
- Lefebvre, H. (1974) *La production de l'espace*, Editions Anthropos, Paris.
Trad. Ita. (1976) *La produzione dello spazio, Volume primo, Volume secondo*, Moizzi Editore, Milano.
- Leroy, A., Mousseau, V., Pirlot, M. (2011) Learning the parameters of a multiple criteria sorting method. *Lecture Notes in Computer Science*, 6992: 219–233.
- Li, Q., Zhang, T., Wang, H., Zeng, Z. (2011) Dynamic accessibility mapping using floating car data: a network-constrained density estimation approach, *Journal of Transport Geography*, 19: 379-393.
- Liao, F. H., Farber, S., Ewing, R. (2014) Compact development and preference heterogeneity in residential location choice behaviour: A latent class analysis, *Urban Studies*, 52(2): 1–24.
- Likert, R. (1932) *Technique for the measure of attitudes Arch. Psycho*, 22(140).
- Lindsey, G., Maraj, M., Kuan, S. (2001) Access, equity, and urban greenways: An exploratory investigation, *Professional Geographer*, 53(3): 332-346.
- Ling, Z., Ni, Y., Cherry, C.R., Li, K. (2014) Pedestrian Level of Service at Signalized Intersections in China Using Contingent Field Survey and Pedestrian Crossing Video Simulation, *Transportation Research Board Annual Meeting*, 14: 41-52.
- Linneker, B., Spence, N. (1996). Road transport infrastructure and regional economic development: the regional development effects of the M25 London orbital motorway, *Journal of Transport Geography*, 4(2): 77-92.
- Liverani, M. (2006) *Uruk la prima città*, Laterza, Roma.
- Loehlin, J. (2004) *Latent variable models, an introduction to Factor, Path and Structural Equation Analysis*, Lawrence Erlbaum Associates Publisher, London.
- Lotfi, S., Koohsari, M.J. (2009) Measuring objective accessibility to neighborhood facilities in the city (A case study: Zone 6 in Tehran, Iran), *Cities*, 26: 133-140.
- Lovetta, A., Haynes, R., Sunnenberga, G., Galeb, S. (2002) Car travel time and accessibility by bus to general practitioner services: a study using patient registers and GIS, *Social Science & Medicine*, 55: 97-111.
- Lucy, W. (1981). Equity and planning for local services. *Journal*, 47(4): 447-457.
- Luo, W., Wang, F. (2003) Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region, *Environment and Planning B*, 30: 865-884.

- Lyon, M., Douglas, G. (2015) *Tactical Urbanism: Short-term Action for Long-term Change*. Island press, Washington, D.C.
- Maslow, A.H. (1954) *Motivation and Personality*. Harper and Row, New York.
- Mateo-Babiano, I. (2016) Pedestrian's needs matter: Examining Manila's walking environment, *Transport Policy*, 45: 107-115.
- Mill, J. S. (1863) *Utilitarianism*. Parker, Son, and Bourn, London.
Disponibile su <https://archive.org/details/a592840000milluoft> (ultimo accesso 30/01/2017).
- Morris, J.M., Dumble, P.L., Wigan, M.R. (1979). Accessibility indicators for transportation planning. *Transportation Research A*, 13: 91-109.
- Muraleetharan, T., Adachi, T., Hagiwara, T., Kagaya, S. (2005) Method to determine pedestrian level-of-service for crosswalks at urban intersections, *Journal of the Eastern Asia Society for Transportation Studies*, 6: 127-136.
- Muraleetharan, T., Adachi, T., Uchida, K., Hagiwara, T., Kagaya, S. (2004) A study on evaluation of pedestrian level of service along sidewalks and at crosswalks using conjoint analysis, *Infrastructure planning review*, 21: 727-735.
- Nakamura, K. (2015) The spatial relationship between pedestrian flows and street characteristics around multiple destinations. *International Association of Traffic and Safety Sciences*, (in press).
- Nasar, J.L., Holloman, C., Abdulkarim, D. (2015) Street characteristics to encourage children to walk, *Transportation Research Part A*, 72: 62–70.
- National Research Council (NRC), *Highway Capacity Manual: HCM2010*. Transportation Research Board, Washington, D.C.
- Neutens, T., Schwanen, T., Witlox, F., De Maeyer, P. (2010) Equity of urban service delivery: a comparison of different accessibility measures, *Environment and Planning A*, 42: 1613-1635.
- Ng, W.Y., Chau, C.K., Powell, G., Leung, T.M. (2015) Preferences for street configuration and street tree planting in urban Hong Kong, *Urban Forestry & Urban Greening*, 14: 30–38.
- Nicholls, S. (2001) Measuring the accessibility and equity of public parks: a case study using GIS, *Managing Leisure*, 6: 201-219.
- Nielsen, T. A. S., Olafsson, A. S., Carstensen, T. A., Skov-Petersen, H. (2013) Environmental correlates of cycling: Evaluating urban form and location effects based on Danish micro-data, *Transportation Research Part D*, 22: 40-44.
- Noriza, R., Ariffin, R., Zahari, R.K. (2013) Perceptions of the Urban Walking Environments, *Social and Behavioral Sciences*, 105: 589 – 597.
- Norušis, M.J. (2011) *IBM SPSS Statistics 19 Advanced Statistical Procedures Companion*, Pearson, London.

- Nussbaum, M. (2011) *Creating Capabilities. The Human Development Approach*, Harvard University Press, Cambridge.
- Nuvolati, G. (1998) *La qualità della vita delle città, Metodi e risultati delle ricerche comparative*, FrancoAngeli, Milano.
- Ogryczak, W. (2000) Inequality measures and equitable approach to location problems, *European Journal of Operational Research*, 122: 374-391.
- Oh, K., Jeong, S. (2007) Assessing the spatial distribution of urban parks using GIS, *Landscape and Urban Planning*, 82(1-2): 25-32.
- Omer, I. (2006) Evaluating accessibility using house-level data: A spatial equity perspective, *Computers, Environment and Urban Systems*, 30: 254-274.
- Pacione, M. (1989) Access to urban services-the case of secondary schools in Glasgow, *Scottish Geographical Magazine*, 105: 12-18.
- Pasaogullari, N., Doratli, N. (2004) Measuring accessibility and utilization of public spaces in Famagusta, *Cities*, 21(3): 225-232.
- Pawlak, Z. (1982) Rough sets, *International Journal of Information & Computer Sciences*, 11: 341-356 .
- Pawlak, Z. (1991) *Rough Set. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Boston/London/Dordrech.
- Pearce, J., Witten, K., Hiscock, R., Blakely, T. (2008) Regional and urban-rural variations in the association of neighbourhood deprivation with community resource access: a national study, *Environment and Planning A*, 40: 2469-2489.
- Piccolo, D. (2010) *Statistica per le decisioni, un'introduzione alla statistica nell'ambiente R*, Il mulino, Bologna.
- Porta, S., Renne, J. L. (2005). Linking urban design to sustainability: formal indicators of social urban sustainability field in Perth, Western Australia. *Urban Design International*, 10(1): 51-64.
- Preston, J., Rajé, F. (2007) Accessibility, mobility and transport-related social exclusion, *Journal of Transport Geography*, 15: 151-160.
- Ramezani, S., Pizzo, B., Deakin, E. (2015) Built Environment versus Personal Traits: an Application of Integrated Choice and Latent Variable Model (ICLV), *14th International Conference on Computers in Urban Planning and Urban Management*.
- Rawls, J. (1971) *A theory of justice*. Harvard University Press, Cambridge.
Trad. Ita. (2008) *Una teoria della giustizia*, Feltrinelli, Milano.
- Recker, W. W., Chen, C., McNally, M.G. (2001) Measuring the impact of efficient household travel decisions on potential travel time savings and accessibility gains, *Transportation Research Part A*, 35: 339-369.

- Robitaille, E., Herjean, P. (2008) An analysis of the accessibility of video lottery terminals: the case of Montréal, *International Journal of Health Geographics*, 7(2).
- Rosero-Bixby, L. (2004) Spatial access to health care in Costa Rica and its equity: a GIS-based study., *Social Science & Medicine*, 58: 1271-1284.
- Roy, B. (1968) Classement et choix en présence de points de vue multiples (la méthode ELECTRE), *La Revue d'Informatique et de Recherche Opérationnelle*, 8: 57–75.
- Roy, B., Bouyssou, D. (1993) *Aide multicritère à la décision: méthodes et cas*, Economica, Paris.
- Saaty, T.L. (1980) *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw Hill, New York.
- Saaty, T.L. (2008) Decision making with the analytic hierarchy process, *Int. J. Services Sciences*, 1(1): 83-98.
- Salonen, M., Toivonen, T. (2013) Modelling travel time in urban networks: comparable measures for private car and public transport, *Journal of Transport Geography*, 31: 143-153.
- Sandel, M. (2010) *Justice, What's the right thing to do?*, Farrar, Straus and Giroux, New York.
- Trad. Ita. (2010) *Giustizia, Il nostro bene comune*, Feltrinelli, Milano.
- Sen, A.K. (1979) Equality of What?, in Matheson M. *The tanner lecture on human values*, University of Utah Press, Salt Lake City.
- Sen, A.K. (1999) *Commodities and Capabilities*, Oxford University Press, Oxford.
- Sen, A.K. (2009) *The idea of justice*, Harvard University Press, Cambridge.
- Trad. Ita. (2014) *L'idea di giustizia*, Mondadori, Milano.
- Singleton, A.D., Longley, P. A., Allen, R., O'Brien, O. (2011) Estimating secondary school catchment areas and the spatial equity of access, *Computers, Environment and Urban Systems*, 35(3): 241-249.
- Smoyer-tomic, K.E., Hewko, J.N., Hodgson, M.J. (2004) Spatial accessibility and equity of playgrounds in Edmonton, Canada, *The Canadian Geographer*, 48: 287-302.
- Sobrie, O., Mousseau, V., Pirlot, M. (2013) Learning a majority rule model from large sets of assignment examples. *Lecture Notes in Computer Science*, 8176: 336–350.
- Soja, E.W. (2010) *Seeking spatial justice*, University of Minnesota Press, Minneapolis.
- Spears, S., Houston, D., Boarnet, M.G. (2013) Illuminating the unseen in transit use: A framework for examining the effect of attitudes and perceptions on travel behavior, *Transportation Research Part A*, 58: 40–53.
- Taleai, M., Sliuzas, R., Flacke, J. (2014) An integrated framework to evaluate the equity of urban public facilities using spatial multi-criteria analysis, *Cities*, 40: 56-69.

- Talen, E. (1997) The social equity of urban service distribution: an exploration of park access in Pueblo, Colorado, and Macon, Georgia. *Urban Geography*, 18: 521-541.
- Talen, E. (2003) Neighborhoods as service providers: a methodology for evaluating pedestrian access, *Environment and Planning B*, 30: 181-200.
- Talen, E., Anselin, L. (1998) Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds, *Environment and Planning A*, 30: 595-613.
- Talu, V. (2013) Qualità della vita urbana e approccio delle capacità, *Archivio di studi urbani e regionali*, 107: 52-73.
- Tanser, F., Gijbetsen, B., Herbst, K. (2006) Modelling and understanding primary health care accessibility and utilization in rural South Africa: An exploration using a geographical information system, *Social Science & Medicine*, 63: 691-705.
- Tempesta, T., Vecchiato, D., Girardi, P. (2014) The landscape benefits of the burial of high voltage power lines: A study in rural areas of Italy, *Landscape and Urban Planning*, 126: 53-64.
- Towne, S. D., Won, J., Lee, S., Ory, M. G., Forjuoh, S. N., Wang, S., Lee, C. (2016). Using Walk Score™ and Neighborhood Perceptions to Assess Walking Among Middle-Aged and Older Adults, *Journal of Community Health*, 41(5): 1-12.
- Truelove, M. (1993) Measurement of spatial equity, *Environment and Planning C*, 11: 19-34.
- Truelove, M. (2000) Services for immigrant women: an evaluation of locations, *The Canadian Geographer*, 44: 135-151.
- Tsou, K.W., Hung, Y.T., Chang, Y.L. (2005) An accessibility-based integrated measure of relative spatial equity in urban public facilities, *Cities*, 22(6): 424-435.
- Van Dyck, D., Cardon, G., Bedorche, B., De Bourdeaudhuij, I. (2011) Do adults like living in high-walkable neighborhoods? Associations of walkability parameters with neighborhood satisfaction and possible mediators, *Health & Place*, 17: 971-977.
- Van Holle, V., Van Cauwenberg, J., Deforche, B., Van de Weghe, N., De Bourdeaudhuij, I., Van Dyck, D. (2015) Do psychosocial factors moderate the association between objective neighborhood walkability and older adults' physical activity?, *Health & Place*, 34: 118-125.
- Vandenbulcke, G., Steenberghen, T., Thomas, I. (2009) Mapping accessibility in Belgium: a tool for land-use and transport planning?, *Journal of Transport Geography*, 17: 39-53.
- West, P.M., Brockett, P.L., Golden, L.L. (1997) A comparative analysis of neural networks and statistical methods for predicting consumer choice, *Marketing Science*, 16(4): 370-391.
- Witten, K., Hiscock, R., Pearce, J., Blakely, T. (2008) Neighborhood access to open spaces and the physical activity of residents: a national study, *Preventive Medicine*, 47: 299-303.

- Yan, A.F., Voorhees, C.C., Clifton, K., Burnier, C. (2010) Erratum to “Do you see what I see?” - Correlates of multidimensional measures of neighborhood types and perceived physical activity-related neighborhood barriers and facilitators for urban youth, *Preventive Medicine*, 50, 18-23.
- Yiannakoulias, N., Bland, W., Svenson, L.W. (2013) Estimating the effect of turn penalties and traffic congestion on measuring spatial accessibility to primary health care, *Applied Geography*, 39: 172-182.
- Yu, W. (1992) ELECTRE TRI: *Aspects méthodologiques et manuel d'utilisation*, Document du LAMSADE No. 74, Université Paris-Dauphine.
- Zadeh, L.A. (1965) Fuzzy sets. *Information Control*, 8: 338–353.