

## APPROFONDIMENTO TECNICHE STATISTICHE

### 1.6 Analisi dei dati

Nella sua configurazione, il questionario consente l'effettuazione di analisi descrittive, bivariate e multivariate. In proposito, è orientamento condiviso in letteratura il trattamento della scala Likert impiegata per valutare la soddisfazione degli intervistati sia a livello complessivo, sia sui singoli aspetti del servizio come una scala a intervallo, quindi con proprietà definibili come "quasi cardinali". Ciò consente il trattamento delle variabili che impiegano le suddette scale anche mediante le tecniche statistiche utilizzabili per le variabili di tipo cardinale.

In secondo luogo, è utile ribadire che la logica sottesa dal modello di rilevazione prevede che la soddisfazione sia considerata come un costrutto intangibile osservabile da due prospettive differenziate:

1. una prima di tipo *non attribute-based*, in virtù della quale si riconosce all'intervistato, grazie alla propria elaborazione cognitiva e affettiva, la capacità di esprimere una valutazione complessiva sulla soddisfazione *overall* nei confronti del servizio; sono infatti presenti, all'inizio e alla fine del questionario, due domande mirate a tale scopo;
2. una seconda di tipo *multi attribute based*, che considera invece la soddisfazione come una variabile multidimensionale formata da diversi elementi costitutivi individuati dai diversi aspetti del servizio, che nel questionario si traducono in altrettante domande<sup>1</sup>.

Dal momento che il questionario deriva da una precedente mappatura del servizio in esame, l'indagine tiene conto di tutti gli elementi del servizio con qualche impatto sul livello di soddisfazione dei fruitori. In questo modo la formazione dell'insieme di indicatori di soddisfazione che si utilizza nell'indagine viene sottratta a valutazioni discrezionali e riduce fortemente il rischio che possano essere esclusi dalla rilevazione aspetti significativi del servizio.

Così progettato, il questionario permette di effettuare analisi statistiche basate sulla ricerca di associazioni e relazioni fra variabili, che sono poi quelle di interesse prevalente per questo tipo di indagine. Il dato di fatto è che risulta possibile avvalersi di uno spettro di tecniche statistiche utilizzabili separatamente ovvero in combinazione per attingere ai rispettivi contributi informativi.

Come si è avuto modo di anticipare nel paragrafo relativo alla fase 4, è bene prevedere un codebook per ogni dataset, dove vengono conservate le informazioni relative alla codifica della variabili. Nella maggioranza dei casi, anche il nome delle variabili (domanda nel questionario) è sostituito da un codice alfanumerico. Per le domande di soddisfazione, questo codice è formato da una lettera (corrispondente alla dimensione a cui la domanda appartiene) e da un numero (che descrive la posizione della domanda in quella dimensione). Anche tale associazione (codice-nome della variabile) deve essere conservata nel codebook.

L'organizzazione del dataset secondo la logica appena indicata ha lo scopo di consentire l'analisi dei dati avvalendosi dei più comuni software per il trattamento dei dati e l'analisi statistica (tra cui ad esempio Excel, Minitab, Spss). Di seguito si riporta l'esempio di un dataset costruito con Excel e del relativo codebook.

---

<sup>1</sup> Gli indicatori di soddisfazione sono raggruppati in 7 dimensioni, in base al criterio dell'omogeneità semantica. Tuttavia, il questionario non contiene domande mirate a rilevare il giudizio di soddisfazione sulla dimensione. Le dimensioni costituiscono un aggregato concettuale che risponde a finalità di analisi, piuttosto che di rilevazione, in quanto hanno il solo scopo di consentire elaborazioni aggregate delle risposte.

	A	B	C	D	E	F	G	H	I	J	K
1	Questionario	P.1	P.2	S.I.	A.1	I.1	...	...	...	S.F.	
2	1	1	1	4	6	5	...	...	...		4
3	2	2	1	3	5	3	...	...	...		4
4	3	4	2	4	6	4	...	...	...		5
5	...	2	1	3	6	5	...	...	...		2
6	...	3	2	2	6	1	...	...	...		4
7	...	1	2	5	3	3	...	...	...		2
8	...	2	2	2	2	1	...	...	...		1
9	n	1	1	6	6	6	...	...	...		5

Figura 1: Esempio di dataset in Excel

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3											
4	Codice	Variabile/domanda		Fasce di età	Codice		Genere	Codice			
5	P.1	Età		18-30	1		M	1			
6	P.2	Genere		31-43	2		F	2			
7	S.I.	Soddisfazione iniziale		44-55	3						
8	A.1	Raggiungibilità del sito		oltre 55	4						
9	I.1	Completezza informazioni									
10	...	...									
11	...	...									
12	...	...									
13	S.F.	Soddisfazione finale									

Figura 2: Esempio di code-book in Excel

Una volta strutturato il dataset, i dati sono pronti per essere elaborati e successivamente interpretati. Le tecniche statistiche, più o meno complesse, utilizzabili per effettuare l'analisi dei dati raccolti sono svariate. La scelta della tecnica statistica da utilizzare deve tenere conto di una serie di fattori quali l'obiettivo dell'analisi e la natura del dato, ossia la tipologia della variabile osservata<sup>2</sup>. Solitamente, è bene partire con un'*analisi uni-variata* dei dati che si occupa di osservare la distribuzione dei caratteri presi singolarmente. Successivamente, può essere interessante passare a un'*analisi bi-variata* e *multivariata* per la ricerca di relazioni fra coppie o gruppi di variabili.

<sup>2</sup> I caratteri statistici possono essere qualitativi – sconnessi (come sesso, religione, luogo di nascita), ordinati (come i giudizi di soddisfazione), quantitativi – discreti (come età e numero di figli) o continui (come peso e altezza). I dati raccolti a seguito dell'applicazione del modello di customer satisfaction on-line sono soprattutto di tipo qualitativo ordinato.

Si precisa che di seguito sono presentati solo i concetti chiave di alcuni strumenti statistici<sup>3</sup>.

### 1.6.1 Analisi univariata

Un primo importante momento dell'analisi dei dati è rappresentato dalle statistiche descrittive (uni-variate) delle variabili considerate singolarmente.

Le statistiche descrittive offrono una rappresentazione iniziale delle principali caratteristiche di ciascuna variabile e consentono di individuare le tecniche statistiche più specifiche da impiegare per una successiva e più approfondita analisi e interpretazione dei dati.

Nel dataset ogni colonna rappresenta l'insieme delle misurazioni, effettuate individuo per individuo, in merito a una determinata caratteristica (distribuzione unitaria). I diversi valori che il carattere può assumere sono detti modalità. Per ottenere una maggiore sintesi è comodo rappresentare la distribuzione di ogni carattere tramite una tabella che riporta la frequenza con cui, nel collettivo di unità in esame, si ripetono le diverse modalità. Questa rappresentazione è nota in statistica come **distribuzione di frequenze assolute**. In pratica, per ogni domanda del questionario, si conta quante volte gli intervistati hanno risposto 1, quante volte 2, quante volte 3 e così via.

Dalla distribuzione di frequenze assolute si può passare alle **frequenze relative**, dividendo le frequenze assolute per il numero totale di osservazioni e a quelle **percentuali** (pari alla frequenza relativa moltiplicata per 100). Ne segue che la somma delle frequenze relative deve sempre essere uguale a 1, mentre quella delle frequenze percentuali deve essere pari a 100. Il vantaggio di usare le frequenze relative o percentuali è apprezzabile quando si vogliono confrontare distribuzioni relative a collettivi di numerosità diversa. Le frequenze assolute, infatti, dipendono dalla numerosità campionaria, mentre quelle relative e percentuali non sono influenzate da tale informazione.

A scopo esemplificativo si supponga di aver misurato la soddisfazione in merito a un servizio su un campione di 10 persone ottenendo i seguenti risultati:

Individuo	Soddisfazione
1	3
2	4
3	3
4	1
5	4
6	2
7	2
8	4
9	5
10	6

**Tabella 1:** Esempio di distribuzione unitaria

La distribuzione di frequenza sarà la seguente:

Punteggio di Soddisfazione	Frequenza Assoluta	Frequenza Relativa	Frequenza percentuale
1	1	$1:10=0,1$	10%
2	2	$2:10=0,2$	20%
3	2	$3:10=0,2$	20%
4	3	$2:10=0,3$	30%
5	1	$1:10=0,1$	10%

<sup>3</sup> Per approfondimenti, teorici e applicativi delle tecniche introdotte, sono disponibili in letteratura svariati testi tra i quali:

- S. Borra, A.Di Ciaccio, (2008), *Statistica: metodologie per le scienze economiche e sociali*, Mc Graw Hill;
- D. Cicchitelli, (2008), *Statistica: Principi e Metodi*, Pearson Education.
- D. Piccolo, (1998), *Statistica*, Il Mulino.

6	1	1:10=0,1	10%
<b>Totale</b>	<b>10</b>	<b>1</b>	<b>100%</b>

**Tabella 2:** Distribuzione di frequenza del carattere soddisfazione

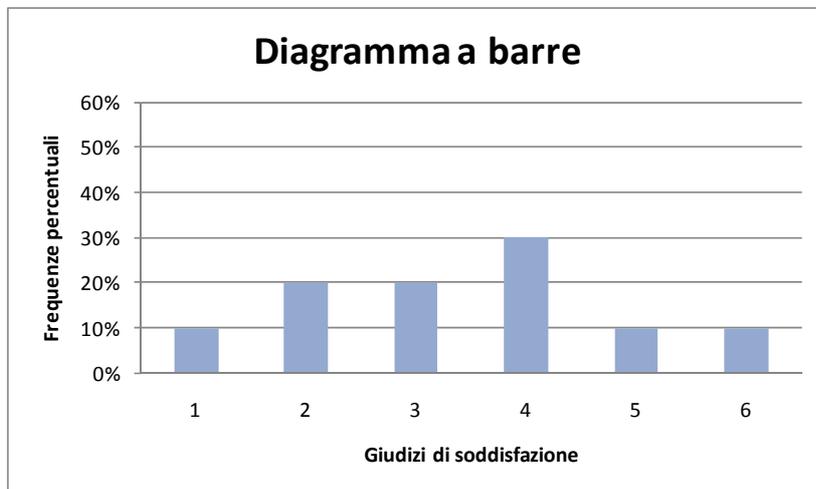
I giudizi di soddisfazione indagati tramite il modello sono misurati con una scala a 6 livelli quindi 6 sono le modalità che il carattere può assumere. La scelta della scala pari permette di accorpare classi di giudizio contigue. Seguendo questa logica i giudizi 1 e 2 possono essere uniti in un unico livello di soddisfazione negativa, 3 e 4 in uno di media soddisfazione e 5 e 6 in uno di alta soddisfazione. Associando a ogni nuova classe un *emoticon*, come nella tabella successiva, è possibile visualizzare immediatamente il giudizio degli utenti.

Classi	Emoticon	Giudizio di soddisfazione
1 -2	☹️	Giudizio negativo
3- 4	😐	Giudizio sufficiente
5- 6	😊	Giudizio positivo

**Tabella 3:** Accorpamento dei giudizi in classi ed attribuzione dell'emoticon

La distribuzione di un carattere può essere rappresentata anche tramite strumenti grafici di facile lettura. Per i caratteri qualitativi ordinati o quantitativi discreti si può utilizzare il **diagramma a barre**, ossia un grafico costituito da tante barre (le colonne) quante sono le modalità che il carattere assume (o le categorie in cui il carattere è stato successivamente diviso). L'altezza di ogni barra è pari alla frequenza (assoluta, relativa o percentuale) della modalità.

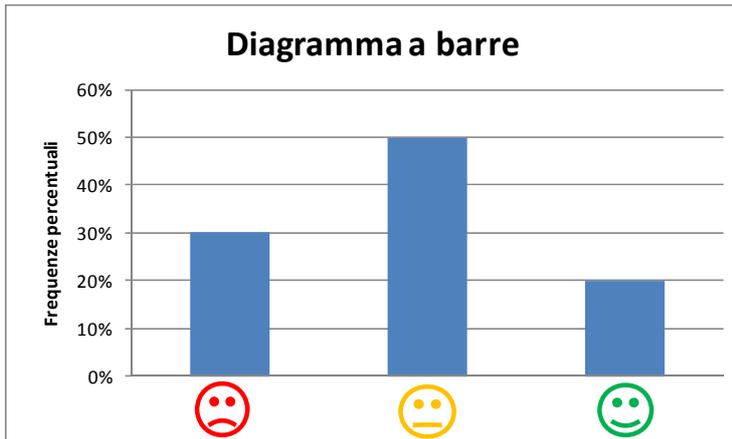
Un esempio di diagramma a barre sulla base dei dati precedenti è il seguente.



**Grafico 1:** Diagramma a barre rappresentativo della distribuzione presa ad esempio

Dalla rappresentazione grafica è agevole osservare che il 10% degli utenti ha dato un giudizio di soddisfazione pari a 1 e il 20% pari a 2. Ne segue che il 30% degli intervistati ha espresso un giudizio negativo in merito al servizio.

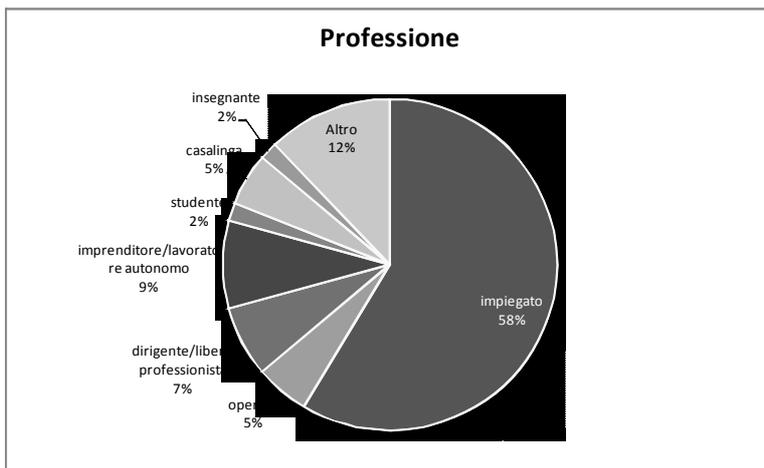
Accorpando le classi con la logica in tabella 3 potremo disegnare il grafico seguente.



**Grafico 2:** Diagramma a barre con classi accorpate e rinominate con gli emoticon

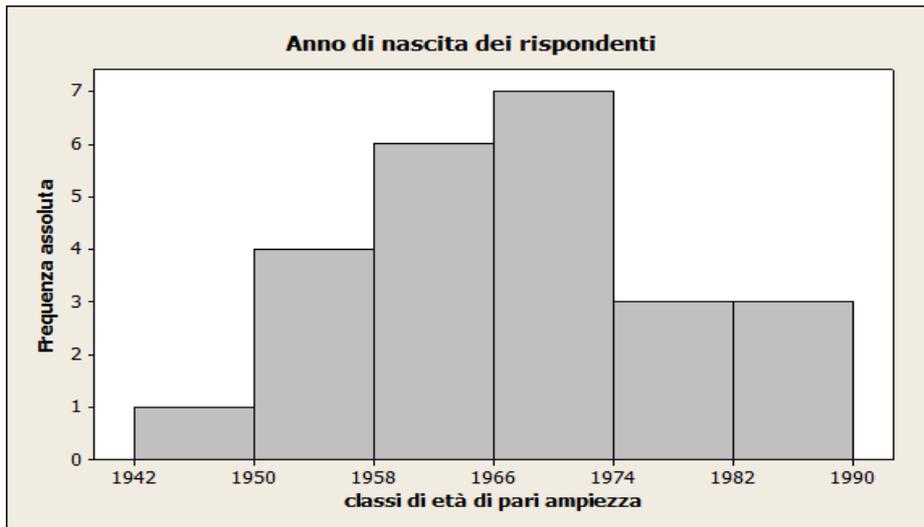
Per la rappresentazione di caratteri non ordinati, invece, sono molto utili i **grafici a torta**. Questo grafico non è altro che una circonferenza divisa in spicchi che rappresentano le diverse modalità che la variabile assume. L'ampiezza di ogni spicchio è proporzionale alla frequenza con cui ogni modalità viene osservata. Questo tipo di grafico è adatto per distribuzioni con un numero di modalità non troppo elevato, al fine di potenziarne la facilità di lettura e interpretazione.

Un esempio di grafico a torta per il carattere è di seguito riportato.



**Grafico 3:** Rappresentazione grafica del carattere professione

Nel caso di un carattere quantitativo con molte categorie o quantitativo continuo, lo strumento grafico adatto per la rappresentazione della distribuzione è l'istogramma. L'istogramma è costituito da tanti rettangoli tra loro contigui quante sono le classi in cui il carattere è diviso. La base delle colonne è pari all'ampiezza della classe (ossia la differenza tra l'estremo superiore e inferiore della classe). Le classi possono essere di ampiezza uguale o diversa. L'altezza delle colonne è pari alla frequenza assoluta della classe se le classi sono di pari ampiezza, e alla densità, ossia la frequenza diviso l'ampiezza della classe, nel caso contrario. Un esempio di istogramma è riportato nella figura sottostante.



**Figura 3:** esempio di istogramma per un carattere diviso in classi di pari ampiezza

Il carattere può essere sintetizzato anche attraverso l'uso di uno o più indici **di posizione e di variabilità**. Alcuni indici di posizione sono:

- la **media aritmetica**, cioè la somma dei valori osservati divisa per il numero di osservazioni. Nell'esempio sopra riportato (tabella 2) la media sarà data da:

$$\text{media} = (3+4+3+1+4+2+2+4+5+6)/10=34/10=3,4$$

È bene ricordare che la media è un indice calcolabile solo per caratteri quantitativi, ma nella prassi viene usato anche per altri caratteri quali ad esempio caratteri come i giudizi di soddisfazione essendo un indice molto facile da calcolare e da interpretare.

- la **mediana**: è un indice più "robusto" rispetto alla media, perché è meno sensibile ai valori estremi/anomali ed è inoltre calcolabile anche per i caratteri qualitativi ordinali. Rappresenta il valore che divide a metà la distribuzione dopo aver ordinato le osservazioni dalla più grande alla più piccola. Nel caso in cui il numero di intervistati (n) sia dispari, la mediana è la modalità presentata dall'unità che, nell'elenco ordinato, occupa la posizione  $(n+1)/2$ ; viceversa se il numero di intervistati è pari, la mediana è il valore medio tra la modalità presentata dall'unità che occupa la posizione  $(n/2)$  e quella che occupa la posizione  $[(n/2)+1]$ . Per esempio, sulla base dei dati sopra riportati l'insieme dei valori osservati ordinati in modo crescente è il seguente:

valori ordinati = (1,2,2,3,3,4,4,4,5,6)

$n=10$  è pari e quindi la mediana è la media tra il valore nella posizione  $(10/2=5)$  che è 3 e il valore nella posizione  $[(10/2)+1=6]$  ossia 4. La media tra 3 e 4 è 3,5. Questo valore identifica la mediana.

Se le osservazioni fossero state  $n=11$ , come nel caso successivo, la mediana sarebbe il valore nella posizione  $[(11+1)/2]$ .

valori ordinati = (1,2,2,3,3,4,4,4,5,6,6)  
mediana = 4.

- la **moda**: può essere calcolata per qualsiasi tipo di carattere e rappresenta la modalità della distribuzione che si presenta con la massima frequenza. Nell'esempio finora proposto la moda è rappresentata dalla modalità 4 che si presenta il maggior numero di volte (3 volte).
- i **percentili**: sono quei valori che dividono la distribuzione in parti di uguale numerosità. In tal senso la mediana è il 50-esimo percentile (Q2). I percentili più utilizzati sono il 25-esimo percentile o

primo quartile (Q1) e il 75-esimo percentile o terzo quartile (Q3), che insieme alla mediana dividono la distribuzione in quattro parti uguali. In particolare tra il primo e il terzo quartile si trova il 50% delle osservazioni.

Altri indici sintetici della distribuzione sono gli indici di variabilità che indicano quanto variano le osservazioni all'interno di uno stesso campione. Tra i numerosi indici di variabilità di un carattere (varianza, deviazione standard, coefficiente di variazione) gli **intervalli di variabilità** sono utilizzabili anche nel caso di caratteri qualitativi ordinati come i giudizi di soddisfazione. Questi indici confrontano due valori caratteristici della distribuzione. Il **range**, per esempio, è pari alla differenza tra il valore più grande e il valore più piccolo dell'insieme dei valori osservati e ordinati in senso crescente.

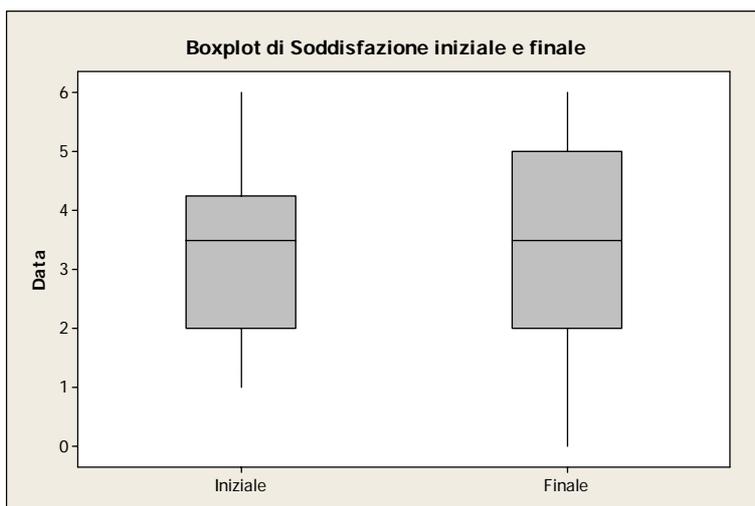
Il cosiddetto riassunto a cinque numeri della distribuzione – l'osservazione più piccola osservata, il primo quartile, la mediana, il terzo quartile e l'osservazione più grande – è uno strumento base per l'elaborazione di altre rappresentazioni grafiche come il **box-plot**. Gli elementi principali che caratterizzano il box-plot sono:

- un rettangolo la cui altezza va dal primo al terzo quartile,
- una linea orizzontale nella scatola che rappresenta la mediana,
- due segmenti che partono dal rettangolo i cui estremi rappresentano il valore minimo e il valore massimo della distribuzione.

L'altezza della scatola rappresenta la distanza interquartile. Questa quantità è un'informazione circa la variabilità delle osservazioni. La posizione della mediana rispetto alla base inferiore (Q1) e superiore (Q3) del rettangolo fornisce informazioni circa la simmetria o asimmetria della distribuzione.

Questo grafico è interamente riassuntivo della distribuzione. Il confronto tra più box-plot permette un confronto sinottico tra le distribuzioni dei caratteri in esame. Affinché il confronto sia appropriato è necessario che i box plot relativi a caratteri diversi (ma misurati nella stessa unità di misura) o a campioni o sottocampioni diversi, vengono rappresentati sul medesimo grafico. In tal modo si evitano distorsioni dovute alla differenza di scala rappresentativa.

Nel grafico sottostante sono riportati due box-plot relativi, rispettivamente, alla soddisfazione iniziale e finale. Dall'esempio si può immediatamente notare che la distribuzione della prima soddisfazione risulta più concentrata. Essa, infatti, presenta una scatola più schiacciata. Inoltre la prima distribuzione presenta una linea orizzontale (mediana) più vicina al terzo quartile che al primo. Questa informazione è sinonimo di asimmetria della distribuzione che, invece, non si riscontra nel caso della seconda distribuzione in cui la mediana è posta circa al centro della scatola.



**Grafico 4:** Rappresentazione grafica tramite box-plot della soddisfazione complessiva iniziale e finale

### 1.6.2 Analisi bivariata

In fase di analisi dei dati è spesso significativo intraprendere lo studio delle relazioni tra due variabili. In letteratura l'analisi congiunta di due caratteri è nota come analisi bivariata. In questa sede useremo il termine di analisi bivariata in senso lato facendo rientrare in questa categoria di strumenti anche tecniche non strettamente statistiche di rappresentazione grafica del contributo integrato di due tipologie di informazioni (per esempio la cosiddetta mappa delle priorità).

Nell'ambito di un'indagine, potrebbe essere interessante, per esempio, analizzare coppie di variabili come la soddisfazione complessiva e il genere (maschio o femmina). Operativamente, per coppie di variabili va inteso l'insieme costituito da due colonne del dataset.

Si consideri un dataset formato dalle variabili soddisfazione complessiva e genere (vedi tabella seguente) misurate su un campione di 20 persone, 12 uomini (codificati con il numero 1) e 8 donne (codificate con il numero 2):

Individuo	Soddisfazione	Genere
1	3	1
2	4	2
3	3	2
4	1	2
5	4	1
6	2	1
7	2	1
8	4	2
9	5	2
10	6	2
11	3	1
12	4	1
13	3	1
14	1	1
15	4	2
16	2	2
17	2	1
18	4	1
19	5	1
20	6	1

**Tabella 4:** Esempio di distribuzione unitaria doppia

L'informazione contenuta nella distribuzione sopra riportata è sintetizzabile in una **tabella di contingenza o a doppia entrata**. In una tabella a doppia entrata le righe della tabella corrispondono alle modalità della variabile riga; viceversa le colonne sono le modalità della variabile riportata in colonna. I valori all'interno della tabella sono l'insieme delle **frequenze congiunte**  $n_{ij}$  ossia il numero di individui (frequenza assoluta) che presentano congiuntamente la modalità  $i$ -esima del primo carattere e  $j$ -esima del secondo carattere. I "totali" di riga e di colonna sono contenuti rispettivamente nell'ultima riga e nell'ultima colonna della tabella e corrispondono alle **distribuzioni marginali** (ossia le distribuzioni semplici) delle due singole variabili<sup>4</sup>.

Nel caso specifico, la tabella a doppia entrata sarà pertanto composta da sei righe (perché la soddisfazione è misurata con una scala a sei punti) e due colonne (una per ogni genere):

---

<sup>4</sup> La frequenza nella cella di una tabella a doppia entrata può essere espressa anche come frequenza relativa se si dividono le frequenze assolute per il numero totale di individui.

Soddisfazione	Genere		Totale
	M=1	F=2	
1	1	1	2
2	3	1	4
3	3	1	4
4	3	3	6
5	1	1	2
6	1	1	2
<b>Totale</b>	<b>12</b>	<b>8</b>	<b>20</b>

**Tabella 5:** esempio di tabella a doppia entrata

E' possibile accorpare le celle (ossia le modalità della soddisfazione) e costruire la tabella con i valori accorpati e rinominati con gli *emoticon*. Si avrà, quindi, una tabella con tre righe e due colonne e in ogni cella si riporterà la frequenza assoluta (congiunta) della nuova categoria derivante dalla somma delle frequenze dei livelli accorpati.

Soddisfazione	Genere		Totale
	M=1	F=2	
	(1+3)=4	(1+1)=2	6
	(3+3)=6	(1+3)=4	10
	(1+1)=2	(1+1)=2	4
<b>Totale</b>	<b>12</b>	<b>8</b>	<b>20</b>

**Tabella 6:** Esempio di tabella a doppia entrata

Mentre le frequenze marginali di colonna non cambiano in quanto i livelli della distribuzione genere non hanno subito modifiche, la distribuzione marginale di riga subisce una variazione per l'effetto dell'accorpamento.

### 1.6.3 Analisi dei giudizi di importanza attribuiti alle dimensioni

Uno degli obiettivi dell'analisi dei dati è quello di contribuire a individuare le aree di miglioramento della qualità del servizio e a fissare delle priorità di intervento.

Il questionario è progettato anche per rispondere a questo scopo. Si ricorda infatti che vengono effettuate tre tipologie di misurazioni:

1. il livello di soddisfazione complessiva nei confronti del servizio: questo dato si rileva all'inizio del questionario per ottenere una valutazione "istintiva" e alla fine, nell'intento di ottenere un giudizio maggiormente ponderato, alla luce delle risposte fornite in precedenza;
2. il livello di soddisfazione dei cittadini intervistati sulle componenti per così dire "elementari" del servizio (cosiddetti indicatori di soddisfazione): queste hanno un riflesso diretto sulla qualità complessiva, ma la loro valutazione puntuale permette una maggiore focalizzazione quando è necessario comprendere in dettaglio su quali aree intervenire;
3. il giudizio di importanza che gli intervistati esprimono sulle dimensioni del servizio; si ricorda che le dimensioni raggruppano elementi puntuali del servizio dal significato omogeneo: quindi per esempio, tutti gli aspetti legati all'informazione (dalla completezza all'aggiornamento) fanno riferimento alla dimensione "adeguatezza delle informazioni". Come si è avuto modo di rappresentare a più riprese in questo documento, "la dimensione può essere assimilata a un concetto più complessivo e per certi versi sintetico della soddisfazione".

Una volta completata l'indagine, si dispone quindi dei dati per calcolare tre grandezze medie: il giudizio di soddisfazione complessiva (misurato in testa e in coda al questionario); il giudizio di soddisfazione sulle singole componenti del servizio; il giudizio di importanza attribuito alle singole dimensioni del servizio medesimo.

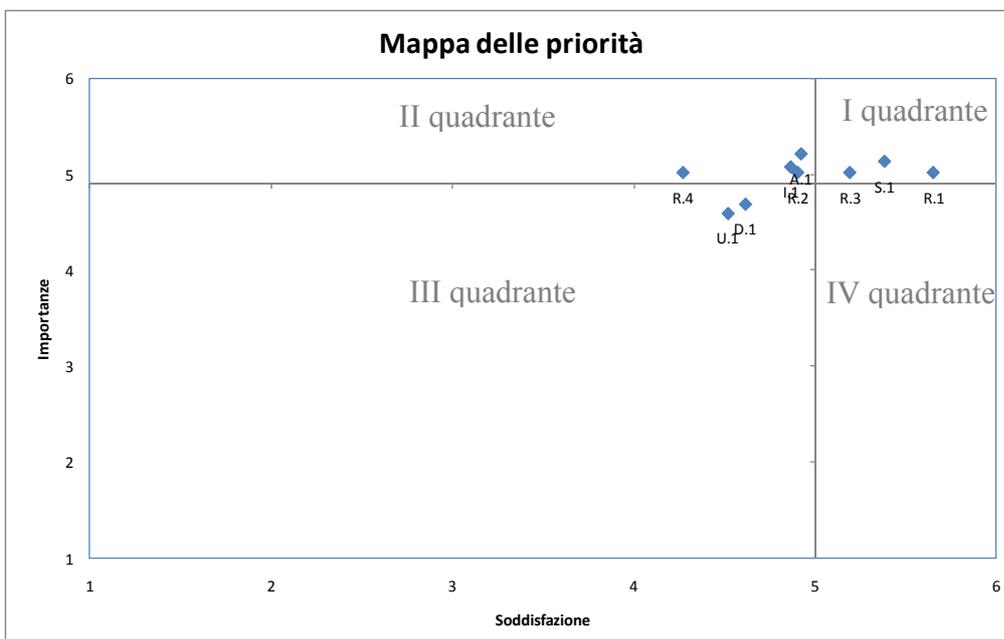
Ai fini della presente analisi, porremo l'accento sui giudizi medi di soddisfazione relativi ai singoli indicatori e ai giudizi medi di importanza delle singole dimensioni.

La logica di fondo cui si fa riferimento è che l'importanza della dimensione per come percepita dai cittadini debba rappresentare il driver per l'individuazione delle priorità di intervento: i giudizi di importanza servono pertanto a facilitare il processo decisionale. Immaginiamo che due indicatori siano giudicati egualmente soddisfacenti; va da sé che dovendo predisporre un'azione migliorativa sulla qualità del servizio, almeno in prima battuta ad essere tenuto in maggiore considerazione dovrà essere quello che appartiene alla dimensione cui i cittadini assegnano una valenza maggiore.

Per definire un primo quadro interpretativo, i giudizi di soddisfazione riferiti ai singoli indicatori di qualità del servizio e i giudizi di importanza attribuiti alle dimensioni possono essere trattati nel modo spiegato di seguito. Un indicatore di soddisfazione va considerato come "prioritario" quando ha un livello di importanza alto ed uno scarso livello di soddisfazione. L'individuazione di indicatori con queste caratteristiche consente al manager di evidenziare le principali aree di miglioramento indirizzando gli interventi verso gli aspetti del servizio ritenuti più importanti e allo stesso tempo meno soddisfacenti dai cittadini.

Questo concetto può essere rappresentato sia in forma grafica, sia tabellare. Nel primo caso si ottiene una rappresentazione sinottica del posizionamento dei diversi indicatori di soddisfazione all'interno di una mappa – cosiddetta "mappa delle priorità" – rappresentata da un piano a due dimensioni e quattro quadranti (vedi figura sotto). Le dimensioni, ossia gli assi della mappa, sono rispettivamente costituiti:

- sull'asse delle ascisse (orizzontale) dal livello di soddisfazione medio (riferito al singolo indicatore di volta in volta considerato), per come rilevato impiegando la scala di misura con intervallo 1-6;
- sull'asse delle ordinate (verticale) dall'importanza media della dimensione cui ogni singolo indicatore appartiene, anch'essa misurata secondo una scala con intervallo 1-6.



**Figura 4:** Esempio di mappa delle priorità

Costruendo le coordinate per ogni singolo indicatore (giudizio medio di soddisfazione e giudizio medio di importanza della dimensione di appartenenza) è possibile riempire lo spazio con una nuvola di punti: ogni punto individua la posizione di un indicatore rispetto agli assi.

Nell'intento di facilitarne ulteriormente la lettura e favorire la presa di decisioni, la mappa viene suddivisa in quattro quadranti. Tra i possibili criteri di *cut-off* dei quadranti si è scelto di utilizzare rispettivamente il valore medio della soddisfazione complessiva finale (quindi quella corrispondente alla domanda di soddisfazione posta a valle del questionario) e la media dei giudizi di importanza di tutte le dimensioni considerate. Il primo divide la mappa verticalmente, il secondo orizzontalmente. Da notare che la superficie

dei singoli quadranti si modifica in funzione dei valori di soddisfazione e importanza e solo in un caso particolare – quello in cui i valori riscontrati di soddisfazione e importanza occupino la posizione centrale dei due assi – sono identiche. Naturalmente, in base ai criteri in esame, ciascuno dei quadranti viene ad assumere un significato preciso, che assume un valore di supporto nel processo decisionale. I quadranti della matrice che vengono così a formarsi esprimono altrettante azioni in ottica di miglioramento, in particolare:

1. Quadrante 1 (Area del monitoraggio): elementi del servizio da controllare nel tempo, per verificare che le aspettative degli utenti non crescano.
2. Quadrante 2 (Area delle criticità): elementi del servizio maggiormente suscettibili di miglioramento, su cui intervenire con priorità più elevata;
3. Quadrante 3 (Area della competitività): componenti che qualificano il valore del servizio e devono essere mantenuti.
4. Quadrante 4 (Area delle illusioni): è possibile che su queste componenti del servizio si stia approfondendo un impegno inutile nell'assicurare agli utenti standard elevati di qualità; si potrebbe considerare di dirottare risorse nell'accrescere il livello di soddisfazione di altri aspetti;

Importanza del singolo indicatore	Giudizio di soddisfazione sul singolo indicatore	
	Basso	Alto
Alta	Area delle criticità Priorità alta (2)	Area della competitività Priorità media (3)
Bassa	Area del monitoraggio Priorità media (1)	Area della illusioni Priorità alta (4)

**Tabella 7:** Il significato dei quadranti nella mappa delle priorità

Parallelamente, si può procedere anche ad una rappresentazione analitica delle informazioni secondo la logica della tabella riportata sotto. Nelle righe della tabella vengono riportati i singoli indicatori di soddisfazione, nelle colonne i valori seguenti:

1. il giudizio medio di soddisfazione per singolo indicatore (S)
2. il giudizio medio di importanza cui il singolo indicatore appartiene (I)
3. il quoziente tra importanza e soddisfazione (I/S).

Domanda del questionario	Soddisfazione media	Importanza media	Quozienti
R.4	4,55	5,46	1,200
U.2	4,41	5,07	1,150
U.4	4,66	5,07	1,088
I.1	5,15	5,48	1,064
R.2	5,15	5,46	1,060
A.1	5,22	5,45	1,044
R.3	5,25	5,46	1,040
U.1	5	5,07	1,014
S.1	5,06	5,13	1,014
PA.1	5,5	5,55	1,009
D.1	5,01	5	0,998
R.1	5,53	5,46	0,987

**Tabella 8:** Quozienti I/S (Importanza/Soddisfazione)

I valori sono ordinati in ordine decrescente, si ottiene una sorta di “classifica” da cui, come nel caso della mappa, è possibile derivare una prima indicazione sulle priorità di intervento.

I due strumenti vanno considerati come integrativi, e non sostitutivi l’uno dell’altro. Dunque in fase di analisi è raccomandabile che vengano costruiti e studiati entrambi per massimizzarne l’apporto informativo, ognuno per le rispettive caratteristiche.

Vediamo perché.

Lo strumento “tabella” ha il vantaggio di consentire, attraverso i quozienti tra importanza e soddisfazione, l’ordinamento degli indicatori. Quanto più alta è la loro posizione nella classifica che viene a determinarsi, tanto più si deve attribuire ad essi una criticità maggiore. Il quoziente tende infatti a crescere in corrispondenza di una importanza elevata e di una insoddisfazione elevata del singolo indicatore.

Tuttavia, si pone un problema di valutazione nell’ipotesi in cui due indicatori ottengano il medesimo quoziente – quindi una identica posizione nella classifica – ma con un dividendo (giudizio di importanza) e un divisore (giudizio di soddisfazione) differenti. Un semplice esempio numerico può chiarire: si supponga che tre indicatori diversi facciano registrare il medesimo quoziente ( $I/S$ ); ciò può avvenire, ad esempio, nel caso proposto: 1)  $I = 6$ ;  $S = 3$ ; 2)  $I = 5$ ;  $S = 2,5$ ; 3)  $I = 4$ ;  $S = 2$ .

Tenendo fede al principio di attribuire al giudizio di importanza il significato di “driver” della criticità di un indicatore, occorrerà stilare una “classifica nella classifica”, facendo occupare una posizione di rango più elevato all’indicatore con importanza maggiore.

Da parte sua, la mappa permette una maggiore visione di insieme nella rappresentazione e facilita l’immediatezza nella lettura. Tuttavia, l’individuazione dei criteri di cut-off – ovvero di delimitazione della superficie dei singoli quadranti della mappa – presenta il limite ineludibile della discrezionalità nella scelta. Nel caso proposto, sono state impiegate rispettivamente la media di soddisfazione complessiva finale e la media di importanza di tutte le dimensioni del servizio. Questa modalità è certamente ragionevole, ma comunque empirica: per esempio, si sarebbe potuto utilizzare la soddisfazione complessiva iniziale, ovvero la media di soddisfazione dei diversi indicatori puntuali. A seconda del criterio adottato, cambiano i confini dei quadranti e conseguentemente può venire a modificarsi il livello di priorità assegnato ai diversi aspetti del servizio. Occorre essere consapevoli che non esiste in proposito un criterio univoco e oggettivo.

Alla luce di quanto appena affermato, è doveroso ribadire che tanto la mappa che la tabella per la fissazione dell’ordine di priorità degli indicatori mediante quoziente  $I/S$ , sono stati concepiti nell’intento di mettere a disposizione dell’amministrazione strumenti di agevole predisposizione che possano fungere da supporto decisionale. Esse vanno pertanto intese come facilitatori e non come mezzi cogenti di decisione che si sostituiscono all’esperienza/competenza del decisore. Gli organismi di governo dell’amministrazione possono – meglio devono – integrare le evidenze emerse dall’indagine alla luce del proprio know-how e degli indirizzi strategici legati alla gestione del servizio.

Alcuni esempi possono essere utili a far comprendere meglio quanto ampio resti lo spazio di manovra dell’amministrazione per intervenire sul miglioramento, integrando col proprio apporto decisionale le evidenze dell’indagine. I criteri metodologici di determinazione delle priorità di intervento, per quanto ragionevoli, non escludono infatti l’adozione di modalità ulteriori che permettano di ottimizzare le logiche di decisione e di ricondurle più efficacemente allo specifico contesto di riferimento.

1. Finora si è detto che la percezione di importanza degli indicatori nell’ottica dei cittadini costituisce il perno concettuale nella determinazione delle priorità di intervento. Pertanto, gli indicatori a importanza elevata e soddisfazione ridotta costituiscono la parte critica su cui intervenire prioritariamente, mentre quelli con importanza e soddisfazione elevate identificano gli aspetti eccellenti del servizio, prefigurando una condizione di capacità che deve essere mantenuta. Invece, l’importanza modesta segnala una priorità minore e, almeno in prima battuta, suggerisce di dirottare l’attenzione altrove, anche nell’ipotesi in cui ad essa corrisponda un livello di soddisfazione modesto.

Tuttavia, nessuno vieta al decisore di mettere mano anche su quegli aspetti del servizio che, per quanto percepiti dai cittadini come scarsamente importanti, hanno comunque riscosso giudizi insoddisfacenti. In proposito possono essere fissati dei criteri empirici per fissare delle condizioni di intervento. Si potrebbe così ipotizzare che tutti gli indicatori di soddisfazione con punteggio minore o uguale a due – sempre rispetto alla scala 1-6 impiegata – possano costituire oggetto di azioni correttive indipendentemente dall’importanza ad essi assegnati. Oppure, secondo lo stesso principio, porre mano a tutti gli indicatori posizionati nel primo quartile della scala di soddisfazione, in modo da considerare l’effettiva distribuzione dei giudizi e non una soglia teorica “assoluta” come nel caso precedente.

2. L'amministrazione potrebbe anche avocarsi la responsabilità di sospendere – persino ignorare – le azioni di intervento su indicatori di criticità elevata. Si pensi ad esempio ad elementi quali la velocità di accesso al servizio o i tempi di erogazione dello stesso: tipicamente su questi temi le aspettative degli utenti risultano sistematicamente elevate e si registra una certa propensione a giudicare insoddisfacente, quasi a prescindere, la qualità del servizio. In realtà i giudizi scontano spesso una qualche asimmetria informativa, in quanto i cittadini esprimono un giudizio per così dire “astratto”, formatosi senza conoscere i meccanismi di funzionamento dell'organizzazione, le cui performance non sempre sono ottimizzabili. A questo occorre anche aggiungere i vincoli di risorse e fattibilità, che spesso rendono impraticabile, anche volendolo, il miglioramento di efficienza nel processo di erogazione del servizio.

Ciò suggerisce che si possa adottare una prospettiva “strategica” nei programmi di miglioramento, che tenga conto della prospettiva interna in aggiunta a quella derivante dalle indagini di soddisfazione. Si creano così le condizioni per simulare diversi scenari di intervento, con la conseguenza di una maggiore flessibilità nella individuazione di percorsi di miglioramento coerenti con le politiche di indirizzo delle singole amministrazioni.

3. Un'ultima notazione: potrebbe anche verificarsi il caso in cui la realtà osservata non presenti criticità di rilievo, dato il consenso unanime riscosso dal servizio oggetto di indagine. La metodologia proposta offre la possibilità di effettuare valutazioni di cambiamento anche in tale contesto: in alternativa alla decisione di mantenimento, potrebbero essere ipotizzati degli scenari di “downsizing” della qualità del servizio – senza naturalmente compromettere la performance complessiva – attraverso la riconfigurazione del processo di erogazione, per esempio mediante riallocazione delle risorse verso ambiti con fabbisogni più impellenti.

Si ripropone quindi la necessità di gestire sinergicamente la conoscenza prodotta su base quantitativa attraverso l'indagine con quella di tipo qualitativo in capo al decisore, impiegando il sistema di analisi come supporto alle decisioni, piuttosto che come “produttore” di decisioni intrinsecamente valide.

#### 1.6.4 Analisi multivariata

Uno degli obiettivi delle indagini di customer satisfaction è la valutazione dell'associazione tra la soddisfazione complessiva e ogni singolo indicatore di soddisfazione. Tuttavia in questo modo si procede a una scomposizione del problema in coppie di variabili e non si ha una visione d'insieme del fenomeno. Esistono in statistica tecniche di analisi più avanzate che permettono di studiare il fenomeno di interesse nella sua interezza, portando alla luce il sistema di relazioni fra variabili: in altre parole si dispone degli strumenti per comprendere come si forma la soddisfazione complessiva nei confronti del servizio a partire dai singoli indicatori di qualità che lo caratterizzano. Tali tecniche sono note come tecniche multivariate di analisi dei dati. Tra gli scopi dell'analisi multivariata, infatti, c'è quello di comprendere la presenza e l'entità degli effetti di interazione tra le variabili. Anche in questo caso la scelta della tecnica da adottare dipende dall'obiettivo che si intende raggiungere e dalla natura del dato. Per studiare ad esempio la dipendenza di un fenomeno da un insieme di variabili si può fare ricorso alla **regressione multipla**<sup>5</sup>. Tramite il modello di regressione multipla si individuano le variabili – nel nostro caso gli indicatori di soddisfazione – che hanno un impatto diretto sul fenomeno di interesse, ossia la soddisfazione complessiva. Tuttavia, un limite dell'analisi di regressione è che la variabile di interesse (la soddisfazione complessiva) può essere influenzata dai fattori che dovrebbero formarla (gli indicatori puntuali di qualità) sia in via diretta che indiretta; inoltre è anche possibile che a loro volta tali fattori interagiscono tra loro. Per ovviare a questo limite metodologico e portare alla luce queste relazioni sono disponibili alcuni modelli statistici noti come **sistemi esperti probabilistici** (PES - *Probabilistic Expert Systems*) o reti bayesiane (BN).

Presentando la regressione multipla e i sistemi esperti probabilistici nei prossimi paragrafi si intende offrire una iniziale rappresentazione delle possibili alternative piuttosto che dare un quadro esaustivo delle tecniche di statistica multivariata. Altri modelli possono essere utilizzati. Tali modelli rispondono a obiettivi molteplici e consentono di ottenere informazioni differenziate che aumentano il know-how del management. Scegliere il modello che risponde in modo coerente agli obiettivi prefissati e la sua corretta applicazione può

---

<sup>5</sup> Il metodo è utilizzabile per variabili continue. Nonostante ciò nella prassi il modello è usato anche per descrivere la soddisfazione complessiva in funzione di una combinazione di indicatori di soddisfazione opportunamente selezionati.

essere complicato per chi non ha competenze specialistiche in questo ambito. Viceversa, l'interpretazione del risultato di un modello statistico è più facile per chi conosce il fenomeno da vicino. Per questo motivo, i modelli statistici in generale e i sistemi esperti in particolare, danno risultati apprezzabili quando sono applicati sotto la supervisione congiunta di statistici ed esperti del problema in oggetto.

### *Regressione lineare multipla*

Con il modello di regressione multipla si intende spiegare un fenomeno – variabile dipendente osservata – tramite le informazioni derivanti da un insieme di variabili dette esplicative, anche esse osservate. Nel modello di regressione multipla la variabile dipendente (codificata spesso con Y) è funzione di una combinazione lineare di variabili esplicative indicate con  $X_1, X_2, \dots, X_n$  opportunamente selezionate.

L'espressione analitica del modello è la seguente:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

I coefficienti  $\beta$  misurano la variazione della variabile dipendente al variare di un'unità della variabile esplicativa ad essi associata ferme restando le altre variabili indipendenti. Il coefficiente di ogni variabile, quindi, è interpretabile come una misura della relazione fra la variabile dipendente e ciascuna indipendente. Il parametro  $\alpha$  indica l'intercetta del modello mentre  $\varepsilon_i$  è un termine di errore.

La selezione delle variabili da inserire nel modello avviene con una serie di procedure statistiche che verificano, mediante opportuni test statistici, la significatività dei coefficienti delle variabili nel modello.

Dal momento che un valore del coefficiente pari a zero annullerebbe la relazione funzionale tra la variabile esplicativa e quella dipendente, per validare la significatività dei coefficienti si verifica l'ipotesi (detta nulla) “il valore del coefficiente è uguale a zero” contro l'ipotesi alternativa “il valore è diverso da zero”. Per decidere quale delle due ipotesi è vera, si osserva un opportuno valore (di probabilità), detto  $p$  (*p value*).

Se questo è superiore al livello di significatività rispetto al quale è stato costruito il test allora si accetta l'ipotesi nulla e si dice che il coefficiente non è significativo, altrimenti si accetta l'ipotesi alternativa. In questo secondo caso è come approvare che il coefficiente è significativamente diverso da zero e quindi la variabile ad esso associata viene inclusa nel modello. Il livello di significatività del test ( $\alpha$ ) è solitamente fissato pari a 0,05.

Il modello di regressione può essere costruito, ossia le variabili esplicative che ne fanno parte possono essere scelte, ad esempio, tramite una procedura per passi detta *stepwise*. La direzione della procedura può essere *forward* o *backward*. Nel caso *forward* si parte stimando un modello che contiene una sola variabile esplicativa e si aggiungono una alla volta le variabili che presentano un coefficiente significativo; nel caso *backward* si parte da un modello che contiene tutte le variabili esplicative e, di volta in volta, si eliminano dal modello le variabili che presentano un coefficiente non significativo. La procedura *stepwise* invece alterna in modo più opportuno passi *forward* e passi *backward*.

Venendo alla realtà di nostro interesse, attraverso il modello si intende spiegare la soddisfazione complessiva – ossia variabile dipendente del modello – tramite gli indicatori di soddisfazione, che svolgono il ruolo di variabili indipendenti (esplicative). Verosimilmente, non tutte le variabili esplicative risulteranno significative nello spiegare la soddisfazione complessiva. La scrematura si esegue sulla base dei risultati dell'output della regressione. La prima informazione dell'output è l'equazione di regressione con tutte le variabili esplicative corrispondenti agli indicatori del questionario, cui si associa la stima dei rispettivi coefficienti che indicano la misura dell'impatto sulla soddisfazione del singolo indicatore. L'output restituisce anche una tabella che riporta per ogni coefficiente (intercetta compresa) informazioni circa la significatività del coefficiente<sup>6</sup>. Di particolare interesse è il valore  $p$ : i valori  $p$  superiori a 0,05 suggeriscono che quella variabile è da eliminare dal modello. Un esempio di equazione di regressione e della tabella con i valori  $p$  associati ai diversi coefficienti è riportato di seguito:

Equazione di regressione calcolata a partire dai dati (giudizi di soddisfazione complessiva verso il servizio e verso i singoli indicatori di qualità) raccolti a seguito di un'indagine di customer satisfaction:

$$\text{Soddisfazione complessiva} = 0,681 + 0,213 A.1 + 0,0998 I.1 + 0,098 R.1 - 0,176 R.2 + 0,200 R.3 + 0,0928 R.4 + 0,0643 D.1 + 0,280 U.1 - 0,0176 U.2 + 0,002 PA.1 + 0,0478 S.1$$

Nella tabella seguente codici e valori associati a ciascun indicatore sono dettagliati per facilità di comprensione.

---

<sup>6</sup> Tale output è ottenuto adottando una procedura *stepwise backward* con il software Minitab. A parità di procedura di selezione l'utilizzo di altri software porterebbe alle medesime conclusioni.

Variabile esplicativa	Descrizione dell'indicatore	Valore del coefficiente di regressione	p value
Intercetta		0,681	0,060
A.1	Riesco ad accedere con facilità al servizio on-line "servizio 0-6"	0,213	0,0025
I.1	Sul sito trovo tutte le informazioni di cui ho bisogno per utilizzare il servizio on-line "servizio 0-6", senza necessità di rivolgermi altrove	0,0998	0,0220
R.1	utilizzare l'"iscrizione 06 on-line" è più vantaggioso rispetto ai canali tradizionali	0,098	0,360
R.2	è facile compilare la domanda di iscrizione 0-6 on-line	- 0,176	0,217
R.3	il tempo di connessione internet concesso per la compilazione della domanda è adeguato	0,200	0,068
R.4	i criteri utilizzati per definire la graduatoria sono di facile comprensione	0,0928	0,0177
D.1	riesco a localizzare le informazioni di cui ho bisogno con pochi click	0,0643	0,456
U.1	questo sito offre la possibilità di contattare facilmente il servizio	0,280	0,021
PA.1	Il servizio on-line "servizio 0-6" semplifica le mie attività	0,002	0,992
S.1	Mi sento sicuro nell'inserire in questo sito dati riservati	0,0478	0,575
U.2	Il sito offre strumenti adeguati per effettuare un reclamo	- 0,0176	0,789

**Tabella 9:** Codice, descrizione, valore del coefficiente e *p value* delle variabili indipendenti (indicatori) del modello di regressione iniziale

Dalla lettura dei valori *p* si riescono a evidenziare le variabili da eliminare dal modello.

Si può procedere ad esempio eliminando una per volta le variabili che presentano coefficienti con valori di *p* superiori a 0,05 partendo dal valore più alto.

L'output che si ottiene alla fine del processo di eliminazione è il seguente:

$$\text{Soddisfazione complessiva} = 1,10 + 0,235 A.1 + 0,200 R.3 + 0,141 R.4 + 0,258 U.1$$

Variabile esplicativa	Descrizione indicatore	Valore del coefficiente	p value
Coefficiente		1,10	0,000
A.1	Riesco ad accedere con facilità al servizio on-line "servizio 0-6"	0,235	0,006
R.3	il tempo di connessione internet concesso per la compilazione della domanda è adeguato	0,200	0,001
R.4	i criteri utilizzati per definire la graduatoria sono di facile comprensione	0,141	0,016
U.1	questo sito offre la possibilità di contattare facilmente il servizio	0,258	0,003

**Tabella 10:** Codice, descrizione, valore del coefficiente e *p value* delle variabili indipendenti (indicatori) del modello di regressione finale

Dall'analisi dei coefficienti ( $\beta_i$ ) si evince che la variabile in grado di contribuire maggiormente a spiegare la soddisfazione complessiva, ossia la variabile che nel modello presenta un coefficiente dal valore più elevato, è la possibilità di contattare facilmente il personale dell'amministrazione (U.1). A seguire le variabili facilità di accesso al servizio on-line (A.1); tempo di connessione per la compilazione della domanda (R.3); facilità di comprensione di criteri per definire la graduatoria (R.4). Considerando la tipologia di variabili esplicative della soddisfazione finale presenti nel modello, si osserva che, sulla base dei dati a disposizione, la soddisfazione del servizio è determinata sia da aspetti legati alla capacità di rispondere alle esigenze dell'utente, sia da aspetti legati all'accessibilità al servizio e alla facilità di contatto.

La bontà di adattamento del modello, ossia quanta variabilità del fenomeno il modello riesce a spiegare a partire dai dati iniziali, è misurabile tramite il cosiddetto coefficiente di determinazione  $R^2$ . Tale valore, espresso in percentuale e compreso fra zero e uno è abitualmente calcolato dal software di analisi<sup>7</sup>. Nell'esempio, un  $R^2$  pari a 83% suggerisce che l'83% della variabilità totale della soddisfazione complessiva è spiegata dalle variabili esplicative conservate (cioè non escluse dalla procedura per mancanza di

<sup>7</sup> Solitamente, si identifica l' $R^2$  con la dicitura R-Sq(adj).

significatività): facilità di accesso al servizio on-line (A.1); tempo di connessione per la compilazione della domanda (R.3); facilità di comprensione di criteri per definire la graduatoria (R.4); possibilità di contattare facilmente il personale dell'amministrazione (U.1). Pertanto, esso sta ad indicare che le variabili esplicative predicono efficacemente la variabile dipendente. In fase di interpretazione di tale valore è opportuno leggere anche l'  $R^2$ -adj (ossia aggiustato) che è il valore con un fattore di penalità per il numero di variabili esplicative. Nell'esempio il valore dell'  $R^2$ -adj raggiunge il valore di 81,7%. Un valore del coefficiente troppo basso suggerisce che il modello di regressione manca di alcune variabili esplicative; viceversa, un valore troppo elevato (maggiore di 0,9) può essere sintomo di correlazione fra le variabili esplicative, come se queste fossero somiglianti l'una all'altra e dunque incapaci di rappresentare in modo appropriato le diverse sfaccettature della realtà osservata. Da notare che il modello di regressione può essere anche usato in modo predittivo per stimare, in un certo tempo T, il valore futuro (non ancora osservato) della variabile indipendente sulla base dei valori assunti dalle variabili indipendenti al tempo T. Per questo motivo le variabili indipendenti sono anche dette predittori.

### *Modelli probabilistici esperti*

Uno dei possibili obiettivi dell'analisi dei dati può essere la comprensione del percorso logico attraverso il quale si genera la soddisfazione dei cittadini in riferimento a un servizio erogato. In pratica si può essere interessati a costruire una "rete" di relazioni dirette e indirette fra la soddisfazione complessiva e gli indicatori di soddisfazione da cui essa è dipendente. Il modello di regressione lineare ammette solo l'individuazione delle variabili direttamente collegate all'effetto soddisfazione complessiva, ma non permette l'individuazione né degli elementi che la influenzano in via indiretta, né delle relazioni che intercorrono tra le variabili. È quindi utile indagare sulla struttura di dipendenza tramite altri strumenti statistici.

Per modellare le relazioni tra le variabili si può far ricorso ai modelli grafici in generale e ai sistemi esperti probabilistici (*Probabilistic Expert System – PES* o reti bayesiane) in particolare. Si tratta di modelli di statistica multivariata per lo studio delle relazioni di dipendenza tra le variabili. Pure se certamente complessi, tali modelli, se consapevolmente adottati, possono rappresentare un valido supporto alle decisioni manageriali<sup>8</sup>. Inoltre, ammettendo una rappresentazione grafica delle relazioni tra le variabili, i sistemi esperti probabilistici sono in grado di facilitare la comunicazione e l'interazione fra decisori con diverso background culturale<sup>9</sup>.

La base della rappresentazione grafica di un PES è una rete (detta grafo), costituita dalle variabili (rappresentate da nodi) e dalle dipendenze fra di esse (rappresentate da frecce). La rappresentazione grafica dei modelli ricorda il diagramma causa-effetto (noto anche sotto il nome di diagramma di Ishikawa), nel quale un effetto – per esempio la soddisfazione complessiva – dipende da una serie di cause dirette e indirette (i singoli indicatori di soddisfazione). Mentre, però, il diagramma causa-effetto ha una natura qualitativa (solitamente, infatti, la sua costruzione deriva da un approccio di tipo brainstorming per l'analisi dei processi) nei PES, la dipendenza/indipendenza tra due variabili (ossia, graficamente, la presenza /assenza tra due nodi), viene studiata direttamente a partire dai dati tramite l'utilizzo di algoritmi computazionalmente efficienti implementati in diversi software per la costruzione di modelli grafici (MIM, Hugin, R). Si osservi che poiché la struttura di relazione viene stimata direttamente dai dati, ossia dalle osservazioni campionarie, se il campione non è statisticamente significativo (estratto casualmente e sufficientemente ampio) è imprudente generalizzare i risultati campionari all'intera popolazione. Questa considerazione è importante anche alla luce dei possibili utilizzi dello strumento.

Una volta sistemato, il modello è a sua volta utilizzabile per simulare scenari secondo la logica "what if" e di conseguenza può essere impiegato come uno strumento di supporto alle decisioni.

Applicato a un dataset relativo a un'indagine di customer satisfaction il modello consente di comprendere dove è più opportuno intervenire sul servizio nell'ottica di migliorarne la qualità.

Un esempio di struttura di dipendenza appresa grazie all'applicazione del modello è il seguente.

---

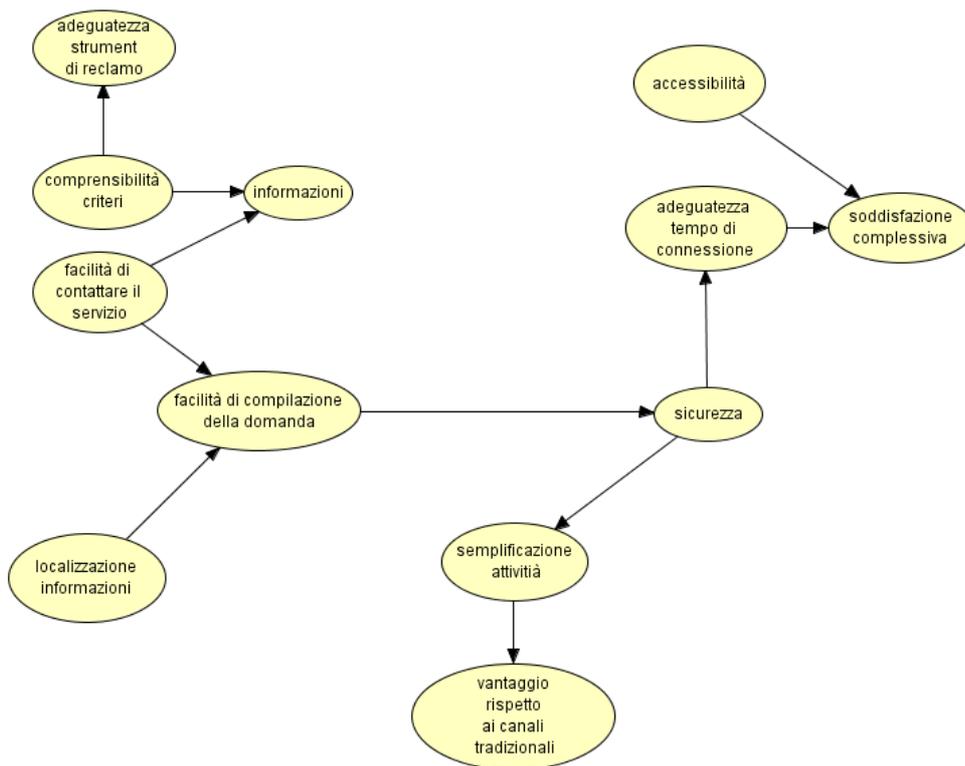
<sup>8</sup> Per approfondimenti si veda: "Probabilistic expert systems for managing information to improve services" di M.F. Renzi, P. Vicard, R. Guglielmetti & F. Musella (2009) in "*The TQM journal*", vol. 21, Issue 4 - pp. 429-442.

<sup>9</sup> Per approfondimenti sul tema dei modelli grafici e dei sistemi esperti probabilistici si veda:

R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, (1999), *Probabilistic Networks and Expert Systems*. Springer, New York;

S.L. Lauritzen, (1996), *Graphical Models*, Clarendon Press;

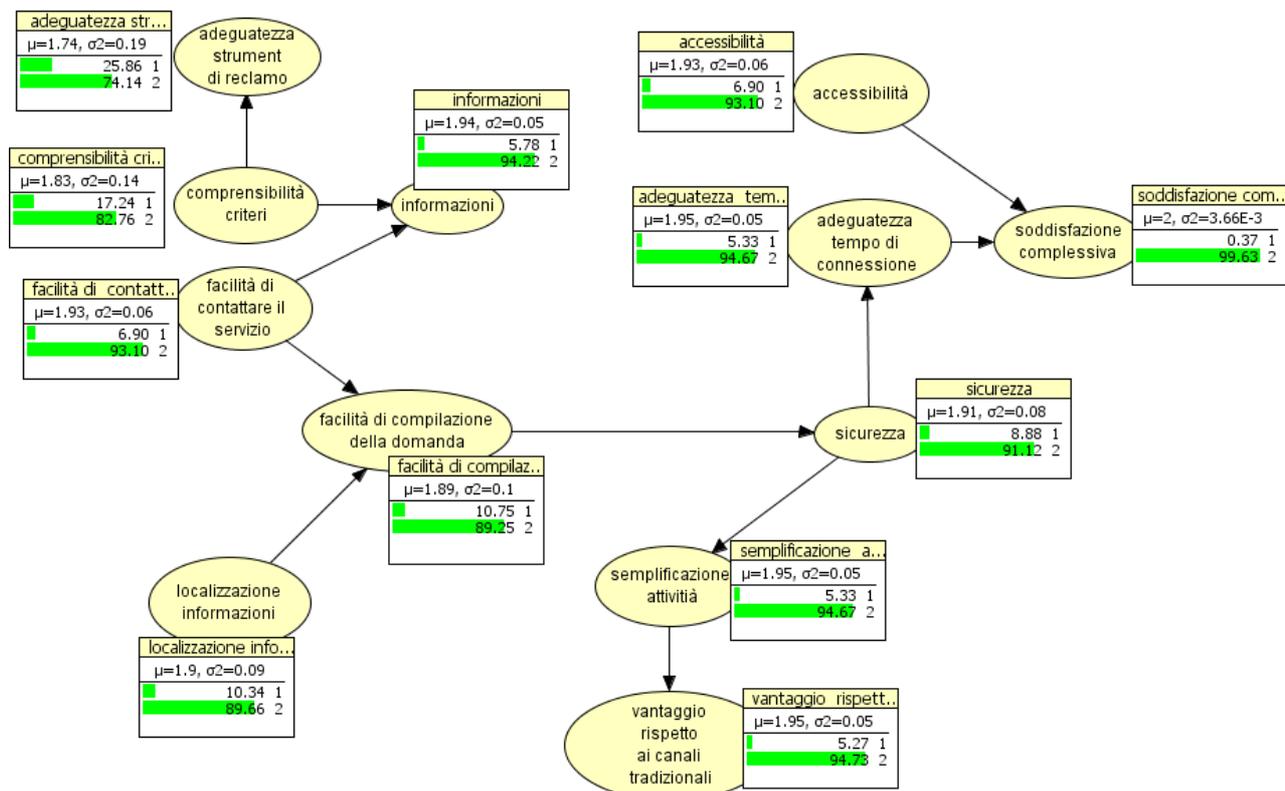
R. E. Neapolitan, (2004), *Learning Bayesian Networks*, Upper Saddle River (NJ), Prentice Hall.



**Figura 5:** Esempio di rete costruita attraverso un sistema esperto probabilistico

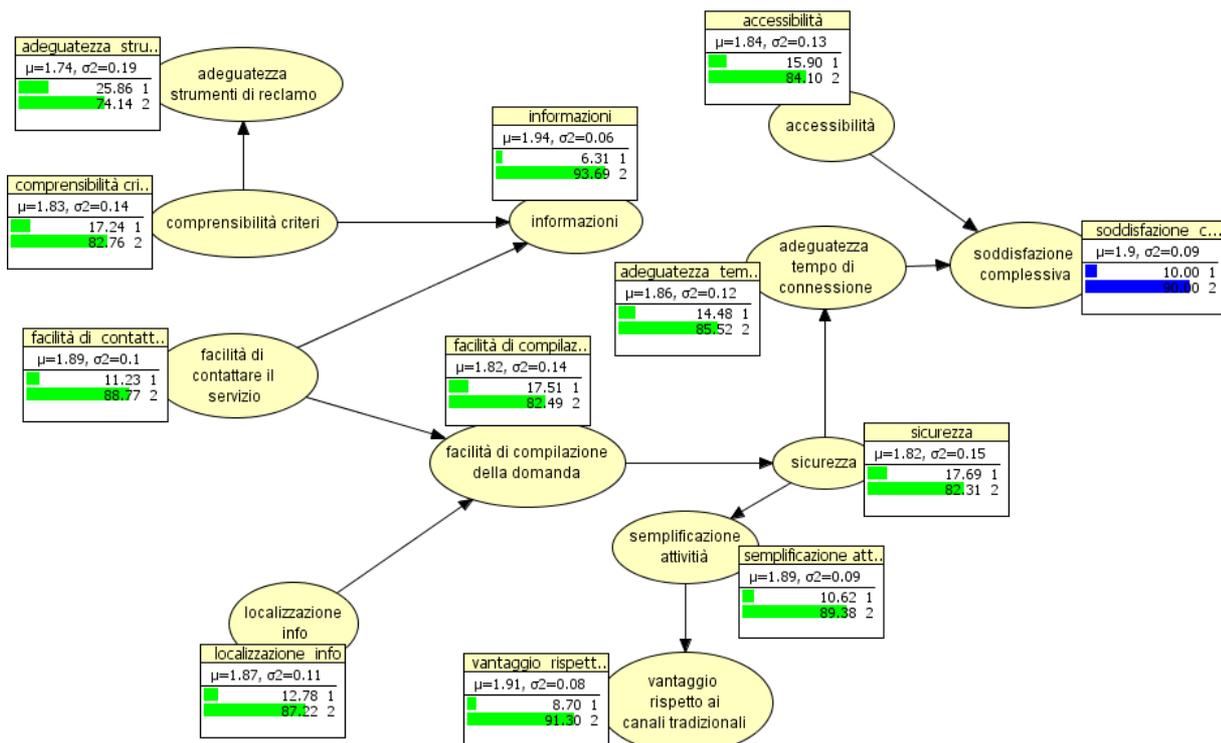
La soddisfazione complessiva è influenzata da due cause dirette (accessibilità al servizio e adeguatezza del tempo di connessione) e da diverse cause indirette.

A ciascuna variabile è associata una tabella di probabilità marginali che mostra la relativa distribuzione (espressa in percentuale) dei giudizi di soddisfazione accorpati, nell'esempio, in due classi (1, 2, 3 primo livello; 4, 5, 6 secondo livello).



**Figura 6:** Probabilità marginale associate ad ogni nodo della rete

E' possibile manipolare la distribuzione marginale di ciascuna variabile e osservare l'impatto (positivo o negativo) che un cambiamento provoca su tutte le restanti. In questo senso il modello si può usare per selezionare le azioni di miglioramento che risultano più efficaci per migliorare il servizio, quando esistono aree di miglioramento. E' possibile, infatti, stimare gli effetti di un'azione mirata a migliorare il giudizio di un singolo aspetto di soddisfazione. Al tempo stesso è possibile valutare quali azioni risulterebbero nocive comportando un abbassamento dei giudizi di soddisfazione. Nell'esempio proposto il servizio sembra riscuotere molti consensi infatti la percentuale di utenti soddisfatta è, per tutte le variabili nel modelli, superiore a quella degli utenti insoddisfatti. In questo caso può essere interessante usare il modello per verificare cosa accadrebbe se si accettasse di innalzare il numero di clienti insoddisfatti fino a raggiungere una soglia fisiologica che non provochi indebolimenti pericolosi nel livello di qualità percepita del servizio.



**Figura 7:** Probabilità marginali modificate dopo un intervento di simulazione

Si osserva in tal caso che per mantenere il livello di soddisfazione complessiva ai livelli desiderati la percentuale di clienti soddisfatti per l'accessibilità può essere ridotta dal 93% all'84% e l'adeguatezza del tempo di connessione dal 95% all'86%. L'intervallo fra punto di partenza iniziale e finale costituisce un margine di tolleranza, all'interno del quale ci si può muovere intervenendo sul processo di erogazione. Nonostante la rappresentazione grafica dei PES renda intuitiva la lettura e l'interpretazione del modello, i PES sono modelli statistici probabilistici complessi per il cui utilizzo è spesso necessario l'ausilio congiunto sia di statistici sia di esperti del fenomeno di interesse. In questo modo si potenziano i vantaggi legati all'utilizzo di uno strumento di questo tipo.