

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE
Corso di Laurea in Statistica e Gestione delle Imprese

**ANALISI SULLA PERCEZIONE DELLA SINDROME
DI BURNOUT DEI PROFESSORI DI EDUCAZIONE FISICA NELLE
SCUOLE SECONDARIE SUPERIORI**

Relatore: Prof.ssa Laura Ventura
Correlatore: Prof. Attilio Carraio

Laureando: Stefano Bragante

ANNO ACCADEMICO 2004/2005

Ringraziamenti

Ringrazio di cuore i miei genitori, la mia ragazza e mio fratello per avermi permesso di raggiungere questo traguardo.

Un ringraziamento speciale alla Prof.ssa Laura Ventura per avermi aiutato di continuo durante lo svolgimento della tesi.

Un doveroso grazie al Prof. Attilio Carraro per il materiale che mi ha fornito e la sua disponibilità.

INDICE DEI CONTENUTI

	pag.
Introduzione	5
Capitolo 1 – Dati e questionario	7
1.1 Il questionario	7
1.2 I dati	9
1.3 Qualità dei dati	10
Capitolo 2	13
2.1 L'analisi preliminare dei dati	13
Capitolo 3	23
Studio dei dati sulla sindrome di burnout tramite l'analisi fattoriale	23
3.1 Metodo di base dell'analisi fattoriale	23
3.2 Comunanza e unicità dei fattori	27
3.3 il processo di estrazione dei fattori	28
3.4 La scelta del numero di fattori da estrarre	30
3.5 I metodi di rotazione dei fattori	32
3.6 Analisi dei dati	33
3.7 analisi cluster applicata al modello stimato con l'analisi fattoriale	42
3.8 Analisi cluster applicata ai dati originari	45
Appendice del capitolo 3	50
Comandi per l'analisi fattoriale	50
Comandi per l'analisi cluster	52
Conclusioni	55
Riferimenti bibliografici	59

INTRODUZIONE

Questa tesi si propone di analizzare come, nella relazione quotidiana e intensa tra un professore e i suoi alunni, questi ultimi possano percepire la condizione di stress cronico, meglio definito come sindrome di burnout che può colpire il loro insegnante.

La sindrome di burnout si manifesta in individui appartenenti a categorie di lavoratori soggetti a periodi prolungati di tensione, ed è caratterizzata da alcuni fattori quali: affaticamento fisico ed emotivo, atteggiamento distaccato e apatico nei rapporti interpersonali, sentimento di frustrazione per mancata realizzazione delle proprie aspettative, perdita della capacità del controllo (smarrimento cioè di quel senso critico che consente di attribuire all'esperienza lavorativa la giusta dimensione). Ciò può far sì che la professione finisca per assumere un'importanza smisurata nell'ambito della vita di relazione e l'individuo non riesce a "staccare" mentalmente tendendo a lasciarsi andare anche a reazioni emotive, impulsive e violente.

In particolare negli insegnanti questa situazione può essere rilevata più facilmente poiché gli studenti trascorrono molto tempo a diretto contatto con loro.

Innanzitutto lo studio si propone di riconoscere se i professori oggetto dell'indagine sono soggetti alla sindrome di burnout; inoltre si aspira ad identificare, tramite un'analisi di classificazione, le eventuali differenze

tra la percezione del burnout di un professore da parte degli studenti delle scuole statali e di quelli delle scuole private su un campione di 193 allievi delle classi terze di alcuni licei della città di Vicenza.

A tal fine è stato predisposto un questionario da somministrare ai due campioni di studenti, nelle medesime condizioni di tempi e modi. Ciò che si vuole verificare è che i casi di burnout sono più frequenti nelle scuole private che non nelle scuole statali. Questa ipotesi è nata da riferimenti bibliografici che riportano testimonianze di studi internazionali svolti in contesti sociali simili più alle scuole private che non alle scuole statali. Si suppone che nelle prime vi sia una più alta possibilità di licenziamento o spostamento in altri settori per gli insegnanti. Tali "paure" danno vita a stress e possono essere causa del cattivo operato di docenza.

CAPITOLO 1

Dati e questionario

1.1 Il questionario

Le domande a cui hanno dovuto rispondere gli allievi delle scuole sono state ricavate da un questionario sulla percezione degli studenti rispetto agli insegnanti affetti da sindrome di burnout, realizzato dal Moshe Tatar e da Vered Yahav dell'Università Hebrew di Gerusalemme, Israele, 1999.

Tale questionario, che è stato tradotto e adattato in modo da poter essere utilizzato in studi anche nelle scuole italiane, si divide in quattro parti:

- nella prima parte sono richieste informazioni riguardo all'allievo rispondente (sesso, età e tipo di scuola: privata o pubblica);
- nella seconda vi sono delle domande aperte riguardanti la descrizione dei comportamenti più salienti degli insegnanti affetti dalla sindrome di burnout;
- nella terza parte, che è quella che è stata utilizzata per questa tesi, viene invece richiesto di rispondere a 14 domande, relative al comportamento del proprio professore, che descrivono con che intensità l'allievo percepisce ciascuna delle caratteristiche tipiche dell'insegnante burnout;

- infine l'ultima parte consiste in 3 domande aperte in cui vengono richiesti pareri per aiutare gli insegnanti in crisi.

La ricerca ha il fine di capire se, secondo gli allievi, i 5 insegnanti di educazione fisica oggetto dell'analisi sono affetti alla sindrome di burnout. Per raggiungere tale obiettivo non sono necessarie la seconda parte e l'ultima parte del questionario che quindi non sono state utilizzate.

Al committente dell'analisi interessa infatti soltanto capire se gli allievi delle scuole private e delle scuole pubbliche, avvertono segni di disagio nei loro professori ed eventualmente conoscere se le impressioni degli alunni corrispondono alla reale condizione dei docenti.

Le 14 domande sono state prese e modificate dalla versione di Hebrew sulla M.B.I. (Maslach Burnout Inventory, Maslach & Jackson 1981).

La traduzione in lingua Italiana del questionario si presenta così:

Classe _____	Età _____	data _____			
Questo questionario serve a studiare il burnout (esaurimento professionale degli insegnanti). Ti chiediamo di rispondere, pensando ad un insegnante di Educazione Fisica che conosci.					
Fino a che punto le seguenti frasi descrivono (si applicano) ad un insegnante burn-out. Usa la seguente scala: 1 corrisponde a mai, 2 a raramente, 3 a talvolta, 4 a frequentemente, 5 a sempre: Un insegnante burnout....					
	1	2	3	4	5
1) Si sente " a terra" alla fine di una giornata intera a scuola					
2) Si sente esaurito alla mattina quando deve andare incontro ad un'altra giornata di lavoro a scuola					
3) Sente che insegnare le/gli dà una grande soddisfazione					
4) Sente che lavorare con gli studenti per un giorno intero è uno sforzo opprimente					
5) Sente che insegnare è frustrante per lei / lui					
6) Sente che lei/ lui deve lavorare troppo duramente nell'insegnamento					
7) Sente che l'insegnamento le/ gli permette di utilizzare al massimo le sue capacità					
8) Sente che lavorando a stretto contatto con gli studenti le /gli crea una grande tensione					
9) Considera di abbandonare l'insegnamento					
10) Sente che insegnare lo sta trasformando in una persona impaziente					
11) Pensa che se potesse ricominciare la sua vita sceglierebbe di nuovo l'insegnamento					

12) Sente che come insegnante si sta realizzando nella vita					
13) Pensa che in un'altra professione (diversa dall'insegnamento) potrebbe meglio utilizzare le sue capacità intellettive					
14) Sente che le aspettative richieste su di lei/lui (dal capo, dal sovrintendente, dal curriculum) sono molto al di sopra delle qualifiche della sua esperienza					

Le risposte sono state inserite in un foglio elettronico: le cui righe rappresentano le unità statistiche e le colonne le diverse risposte possibili per ciascuna domanda. Nella cella corrispondente alla risposta data si inseriva un numero da 1 a 4, a seconda del sesso e della scuola di provenienza. Nell'importazione dei dati nel software statistico open-source R, di cui si è usata la versione 2.1.0, questi sono stati successivamente ricodificati: ciascuna variabile conteneva la risposta a una domanda, inoltre sono state inserite due variabili, relative a sesso e scuola di provenienza.

1.2 I dati

I ragazzi che hanno preso parte alla ricerca sono 193 alunni di classe terza, dell'età di 16 o 17 anni.

Dato che sono interessato a capire se e in che misura la percezione del burnout da parte degli alunni varia in base tipo di scuola frequentata e al sesso, si sono identificate 4 categorie che compongono il campione come illustrato nella seguente tabella.

CATEGORIA	FREQUENZA	PERCENTUALE
Maschi di scuola privata	48	25%
Femmine di scuola privata	29	15%
Maschi di scuola pubblica	20	10%

Femmine di scuola pubblica	96	50%
----------------------------	----	-----

In totale 116 allievi (il 60% dei totali 193) provengono da scuola pubblica e 77 (cioè il restante 40%) da scuola privata.

Confrontando il sottocampione composto dagli allievi delle scuole pubbliche e quello degli allievi delle scuole private si nota che le unità della scuola pubblica sono soprattutto femmine.

Nelle scuole private, invece, la situazione è quasi opposta: infatti sono più i ragazzi che non le ragazze a rispondere e si dispone di un basso numero di unità di sesso femminile.

1.3 Qualità dei dati

Le modalità di risposta misurano l'intensità con cui, secondo gli allievi, il professore avverte la presenza di ciascuno dei 14 aspetti nell'insegnante di educazione fisica, secondo la scala Likert (cfr. Likert, 1932) riprodotta in seguito.

1	2	3	4	5
Mai	Raramente	Talvolta	Frequentemente	Sempre

La scala Likert prevede che una lista di affermazioni, collegate agli atteggiamenti su cui si vuole indagare, venga sottoposta ai rispondenti assieme a 5 possibili alternative di risposta.

Si suppone che la distanza tra le modalità "Frequentemente" e "Sempre" sia la stessa che intercorre tra "Mai" e "Raramente", così come rispetto alla modalità "Talvolta".

Non vi è nemmeno un dato mancante nel dataset: questo aspetto può essere dovuto a diverse cause:

- le domande del questionario sono formulate in modo particolarmente semplice da comprendere: quindi è probabile che i rispondenti non abbiano avuto nessuna difficoltà nel capirne il senso;
- gli allievi possono aver dimostrato interesse per l'argomento dell'indagine, presumibilmente perché le domande sono inerenti ad una questione che li riguarda da vicino, cioè la salute del loro professore di educazione fisica. E' probabile che un buon lavoro di motivazione dei rispondenti sia stato svolto da chi ha seguito l'indagine sul campo.

CAPITOLO 2

2.1 L'analisi preliminare dei dati

Prima di inoltrarsi nell'applicazione di metodi di studio complessi, quali l'*analisi cluster* e l'*analisi fattoriale*, è utile fare delle analisi preliminari dei dati, di tipo descrittivo. Lo scopo di queste procedure è di identificare le più rilevanti caratteristiche dei dati e di capire, tramite l'osservazione di appositi grafici e il calcolo di *test non parametrici*, quali siano le eventuali relazioni tra le variabili. In questo modo è possibile formulare delle ipotesi iniziali inerenti i dati che vengono verificate nelle analisi successive.

Come preannunciato il lavoro è stato eseguito sia in scuole private che in scuole statali di uguale ordine e grado. E' stata scelta come riferimento la classe terza del liceo classico, scientifico, linguistico e motorio, come riferimento istituzionale più tradizionale.

Gli istituti coinvolti sono stati due scuole statali e due scuole private della città di Vicenza, mentre gli insegnanti coinvolti sono cinque, di cui due maschi.

L'ambiente di lavoro è un fattore che gioca un ruolo importantissimo nella vita professionale di un individuo e lo stato di salute mentale di professionisti come i docenti, che lavorano sempre a contatto con ragazzi giovani, non deve essere trascurato.

Come detto in precedenza, interessa capire, relativamente ai professori di cui si dispongono i dati, se la sindrome di burnout si manifesta nelle diverse realtà lavorative delle scuole private e delle scuole pubbliche e con quali differenze.

Ad esempio può essere che nelle scuole private i professori si sentano più soddisfatti o meno soddisfatti del loro lavoro che non quelli che lavorano nelle scuole pubbliche.

La tabella che segue permette di confrontare 3 indici di posizione (1° quartile, mediana e 3° quartile) tra i due tipi di scuola per ogni variabile del questionario.

	Scuola pubblica			Scuola privata		
	1°quar.	Med	2°quar.	1°quar.	Med	2°quar.
D1	2	3	4	3	3	4
D2	2	2	3	2	2	3
D3	3	3	4	3	4	4
D4	2	2	3	1	2	3
D5	1	2	3	1	2	2
D6	2	2	3	1	2	3
D7	2	3	4	3	4	4
D8	1	2	3	1	2	3
D9	1	1	2	1	1	1
D10	1	2	3	1	2	3
D11	2	3	4	2	3	5
D12	2	3	4	3	4	5
D13	2	3	3	2	3	3
D14	1	2	3	1	1	3

La mediana è un indice che dà un'idea dell'opinione media diffusa riguardo a ciascuno degli aspetti abbinati alle variabili del questionario.

Se per una variabile la mediana è 4, ad esempio, significa che la maggior parte delle risposte date dai rispondenti assumano valori elevati.

Con un rapido confronto tra le mediane dei due gruppi analizzati si osserva che soltanto nel sottocampione della scuola privata ci sono mediane pari a 4, in corrispondenza di 3 delle 4 domande che richiedono di esprimere un giudizio sui comportamenti che sono indice di benessere del docente. Nel sottocampione della scuola pubblica le mediane relative alle stesse domande sono più piccole. Questa osservazione induce ad ipotizzare che i professori delle scuole private siano più felici del loro lavoro che non quelli delle scuole pubbliche: nelle analisi successive si verificherà questo aspetto.

La differenza tra il 3° e il 1° quartile viene chiamata *scarto interquartile*, ed è un buon indicatore di quanto è grande la variabilità delle variabili. Ad esempio, se lo scarto interquartile, per una variabile, fosse 0 (come nel caso della variabile d9 nel sottocampione della scuola privata) significherebbe che almeno la metà dei rispondenti hanno risposto col valore mediano: ciò può voler dire che le opinioni dei rispondenti, riguardo a quel determinato aspetto dell'indagine, coincidono nella maggioranza dei casi.

Il caso opposto è quello in cui si ha uno scarto interquartile grande (un esempio riferito al nostro caso è la variabile d11 nel sottocampione della scuola privata, che ha scarto interquartile uguale a 3), che identifica risposte piuttosto eterogenee.

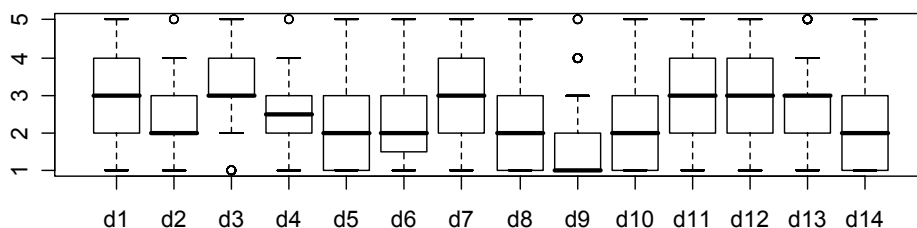
Il 1° e il 3° quartile sono quindi due indici di posizione molto utili per capire quanto omogenee o quanto eterogenee siano le risposte date alle domande all'interno del campione.

Uno strumento grafico che permette di apprendere in modo immediato quali siano le modalità di risposta più gettonate alle domande dagli allievi è il *boxplot*.

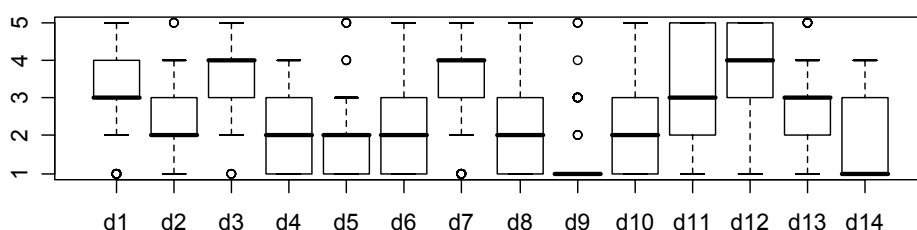
Il grafico **box and whiskers plot**, più comunemente detto **box plot**, è una tipologia di rappresentazione grafica che ricava i dati significativi (i principali quantili) trascurando quelli non importanti. Il grafico è costruito nel modo seguente: si devono calcolare i tre quartili della distribuzione: su un asse orientato, si individua un rettangolo (*scatola/box*) le cui basi orizzontali rappresentano, rispettivamente, il primo e il terzo quartile; la sua altezza, quindi, rappresenta la differenza interquartile; un segmento interno alla scatola identifica la posizione della mediana; esternamente alla scatola, vi sono i 2 *baffi* (*whiskers*) che rappresentano i valori massimo e minimo della distribuzione. Per una distribuzione di frequenza il grafico in questione consente di evidenziare la misura della **dispersione**, la presenza di eventuali **valori anomali** (outliers) e la sua **simmetria o asimmetria**.

Realizzando i boxplot delle risposte alle 14 domande del questionario, rispettivamente per gli allievi delle scuole pubbliche e per gli allievi delle scuole private, si osserva se i sintomi del burnout si ritrovano più nei professori delle scuole pubbliche o in quelli delle scuole private.

Risposte al questionario nelle scuole pubbliche



Risposte al questionario nelle scuole private



Osservando i due insiemi di grafici non si notano differenze di particolare rilevanza tra i due casi: i ragazzi dell'uno e dell'altro gruppo hanno dato risposte più o meno simili senza far trapelare ripetutamente i sintomi della sindrome di burnout.

Infatti almeno il 75% degli allievi, in entrambe le realtà di scuola, hanno mostrato di non notare con regolarità i sintomi del burnout nel comportamento degli insegnanti rispondendo per lo più "mai", "raramente" o al massimo "talvolta" alle domande che considerano la frequenza dei modi di agire classici dell'insegnante stressato (cioè le domande: 1, 2, 4, 5, 6, 8, 9, 10, 13, 14).

I boxplot relativi ad alcune di queste domande evidenziano che nelle scuole pubbliche gli allievi riscontrano le avvisaglie del burnout meno di rado che non nelle scuole private. Per verificare se mediamente vi sono

effettivamente delle differenze sostanziali si può far uso del **test di Wilcoxon-Mann-Whitney**, noto pure come **test U di Mann-Whitney** o **test di Wilcoxon**: questo è uno dei più potenti test non parametrici per verificare, in presenza di valori ordinali, se due campioni statistici provengono dalla stessa popolazione, in particolare serve a constatare l'uguaglianza o la diversità tra i parametri di locazione (le mediane).

L'*ipotesi nulla* (H_0) che voglio verificare tramite questo test è l'uguaglianza tra le mediane dei due campioni (μ_s e μ_p).

L'*ipotesi alternativa* (H_1), considerata nel nostro caso, (cioè quella che viene considerata attinente alla realtà nel caso non fosse veritiera l'ipotesi nulla) è che la mediana per il campione “scuola pubblica” è più grande di quella del campione “scuola privata”.

In letteratura le ipotesi vengono indicate così:

$$H_0: \mu_s = \mu_p \quad \Leftrightarrow \quad H_0: \mu_s - \mu_p = 0$$

$$H_1: \mu_s > \mu_p \quad \Leftrightarrow \quad H_1: \mu_s - \mu_p > 0$$

Caratteristiche del docente	p-value osservato
Stanchezza causata da una giornata di lavoro (d1)	0.9331
Esaurimento prima di iniziare il lavoro (d2)	0.1976
Sforzo dovuto al lavoro (d4)	0.1427
Senso di frustrazione (d5)	0.08388
Mancanza di energie (d6)	0.4419
Tensione per il contatto con gli alunni (d8)	0.07264
Desiderio di smettere di insegnare (d9)	0.0001
Insofferenza dovuta alle fatica di insegnare (d10)	0.01468
Aspirazione ad un lavoro più appagante (d13)	0.1833
Mancanza di capacità di insegnamento (d14)	0.02943

I p-value evidenziati, relativi agli aspetti specificati nella tabella, sono minori di 0,1 e in alcuni casi sono addirittura minori di 0,5.

Come si era supposto ci sono degli items del questionario che rivelano come nelle scuole pubbliche prese in analisi i professori siano più spesso soggetti ai sintomi dello stress che non nelle scuole private.

Anche se è emerso chiaramente che i professori delle scuole statali non manifestano in maniera preoccupante i sintomi della sindrome di burnout, col test di Wilcoxon ho appurato che essi sono più frustrati rispetto ai docenti delle scuole private a causa del lavoro e talvolta sopportano poco di dover insegnare.

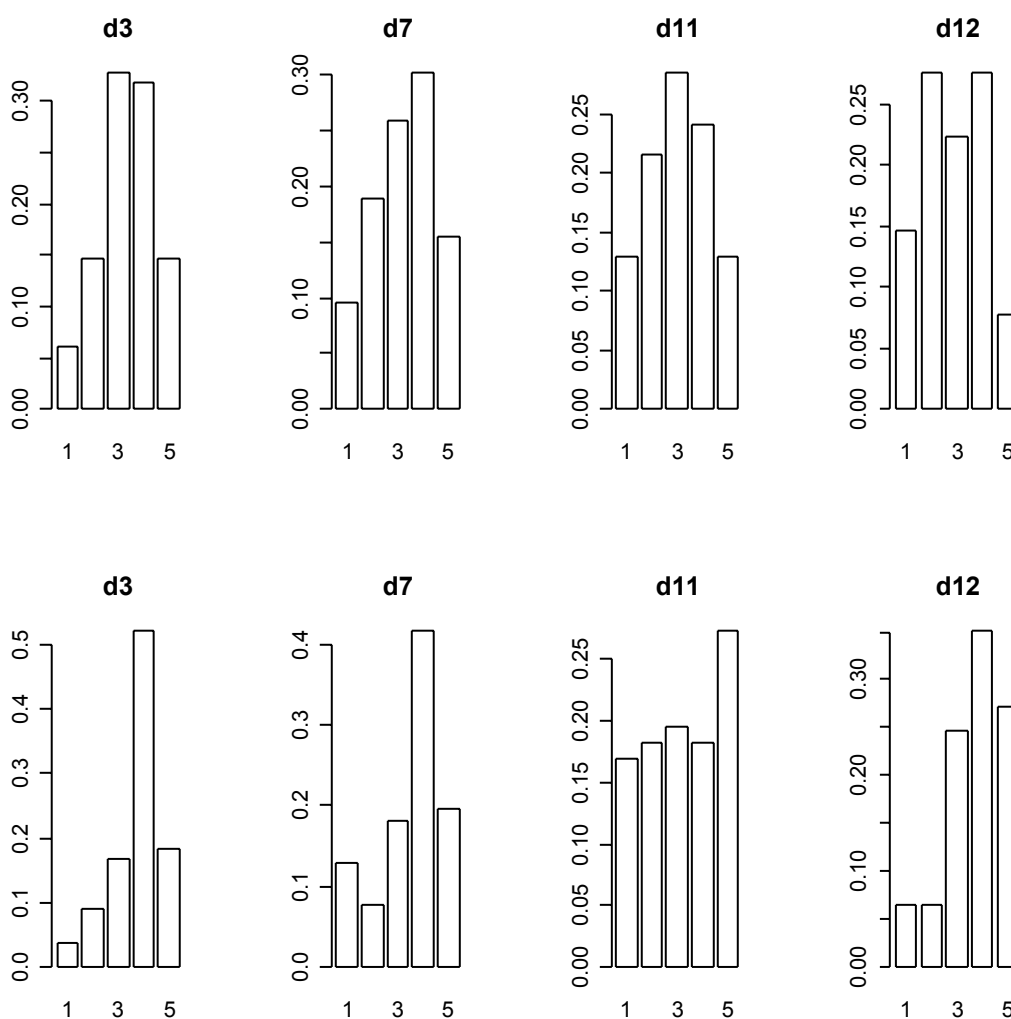
Mentre secondo quasi tutti gli allievi delle scuole private (fanno eccezione pochissimi casi) l'insegnante di educazione fisica non desidera mai cambiare lavoro, nelle scuole statali questa volontà è più manifesta anche se piuttosto di rado.

Tutto sommato, anche se è provato che nelle scuole private la situazione è la migliore, non ci sono elementi dell'analisi che possano far preoccupare sulla salute dei docenti sotto analisi.

D'altro canto le mediane delle risposte alle restanti domande, che misurano la frequenza dei comportamenti che sono indice di benessere del docente, assumono valori non inferiori a 3 indicando vigoria del docente. In particolare, a conferma di quanto precedentemente affermato, sembra che i professori delle scuole private siano più appagati dal loro lavoro che non gli insegnanti delle scuole pubbliche: alle domande "Sente che insegnare le/gli dà una grande soddisfazione", "Sente che l'insegnamento le/gli permette di utilizzare al massimo le sue capacità", "Pensa che se potesse ricominciare la sua vita sceglierebbe di nuovo l'insegnamento", "Sente che come insegnante si sta realizzando nella vita" (domande 3, 7, 11, 12), gli allievi delle scuole private hanno dato risposte elevate con più assiduità.

I seguenti diagrammi a barre mostrano, per ognuna di queste domande, la percentuale di alunni che hanno risposto con ciascuna modalità.

La prima riga di grafici fa riferimento ai questionari compilati dagli allievi delle scuole pubbliche; la seconda riga fa riferimento a quelli compilati dagli alunni delle scuole private.



La tabella che segue mostra i p-value del test di Wilcoxon utilizzato questa volta per controllare se le mediane del campione degli alunni frequentanti la scuola pubblica sono uguali o minori di quelle del campione degli alunni frequentanti la scuola privata.

In altre parole si vuol verificare che la situazione dei docenti della scuola privata sia migliore.

Caratteristiche del docente	p-value osservato
Soddisfazione tratta dall'insegnamento (d3)	0.0049
Utilizzo delle capacità (d7)	0.0594
Risceglierebbe l'insegnamento come professione (d11)	0.1664
Realizzazione di sé stesso/a tramite l'insegnamento (d12)	0.0000

Osservando e confrontando i grafici si osserva che nelle scuole private professori di educazione fisica diano agli alunni l'impressione di trarre molta soddisfazione dall'insegnamento (vedi i 2 grafici d3). Circa il 70% degli allievi infatti ha risposto "spesso" o "sempre" alla domanda sulla soddisfazione del professore. Osservando anche la tabella dei p-value si nota che diversa, ma non sfavorevole, è la situazione dei professori delle scuole pubbliche oggetti dell'analisi: pure essendo più numerosi gli alunni che ritengono che il proprio professore sia contento (circa il 45% degli alunni hanno affermato che il professore di educazione fisica è "spesso" o "sempre" soddisfatto del proprio lavoro) molti non si sbilanciano e sostengono che il docente è "talvolta" appagato dalla propria attività.

La situazione dei professori militanti nelle scuole private è preferibile a quella dei docenti delle scuole pubbliche, come è evidente dall'osservazione dei p-value del test di Wilcoxon, che prova che le mediane delle risposte alle domande 3, 7, 11, 12 è maggiore nel sottocampione "scuola privata".

Molto netta è la differenza tra le risposte date alla dodicesima domanda del questionario dai due gruppi di alunni: il p-value del test è approssimabile allo 0.

Infatti si osserva dai boxplot, ma ancora meglio dai diagrammi a barre, che alla dodicesima domanda (“Sente che come insegnante si sta realizzando nella vita”) gli allievi delle scuole private hanno risposto con punteggi alti nella maggioranza dei casi; quasi mai hanno dato risposte negative.

Molti allievi delle scuole pubbliche (quasi il 50%) invece hanno manifestato di percepire che il proprio docente di educazione fisica è inappagato dal proprio lavoro.

CAPITOLO 3

Studio dei dati sulla sindrome di burnout tramite l'analisi fattoriale

L'**analisi fattoriale** è un metodo statistico che trova larga applicazione nella ricerca psicologica e sociale: essa infatti viene spesso utilizzata per formalizzare teorie nell'ambito di test mentali e del comportamento umano.

Questo tipo di analisi è impiegata per lo studio delle relazioni in un insieme di variabili allo scopo di

- ridurre l'informazione che lo studioso sta ricercando nei dati;
- semplificare la comprensione di un fenomeno;
- individuare delle dimensioni latenti a partire dalle variabili osservate.

Per uno studio approfondito dell'analisi fattoriale e delle sue applicazioni si rimanda a Fabbris (1990).

3.1 Modello di base dell'analisi fattoriale

Partendo da una serie di variabili osservate direttamente sul campione che si sta esaminando si tratta di spiegare perché esse correlano tra di loro, facendo riferimento ad altre variabili, dette *variabili latenti* o *fattori comuni*, che non vengono osservate direttamente ma sono ricavate dal ricercatore, a partire dalle variabili osservate, per spiegare il perché certi fenomeni covariano tra di loro.

Il punto di partenza dell'analisi fattoriale è la matrice di correlazione tra le variabili osservate, denominata R.

Volendo riportare un esempio di 8 variabili, la matrice R è:

	V1	...	V8
V1	1	Cor(v1,v2)	Cor(v1,v8)
V2	Cor(v1,v2)	...	Cor(v2,v8)
...
V8	Cor(v1,v8)	...	1

Il punto di arrivo, invece, è la matrice delle *saturazioni* o *pesi fattoriali*, che si presenta come quella riprodotta qui sotto. Le saturazioni sono indicate con a_{ij} e riferiscono le relazioni che intercorrono tra le variabili osservate V_i ($i = 1, 2, \dots, 8$) e le nuove variabili o *fattori comuni* F_j ($j = 1, \dots, q$, dove q è il numero di fattori latenti estratti), che sono state generate col fine di spiegare perché le variabili correlano tra di loro. Supponendo di aver estratto 2 variabili latenti si avrà la seguente matrice dei pesi fattoriali:

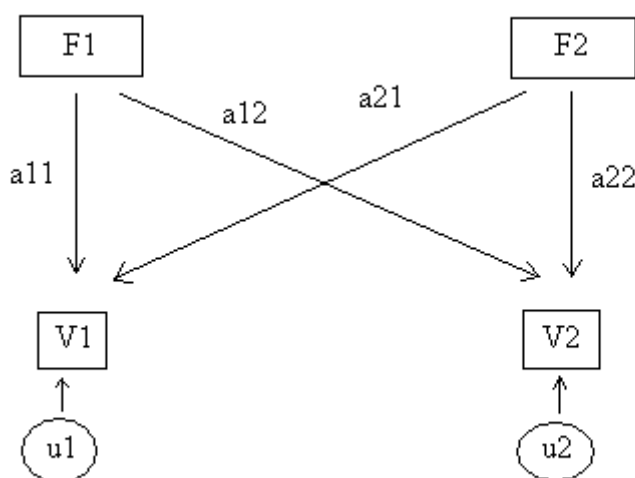
	F1	F2
V1	a_{11}	a_{12}
V2	a_{21}	a_{22}
V3	a_{31}	a_{32}
V4	a_{41}	a_{42}
V5	a_{51}	a_{52}
V6	a_{61}	a_{62}
V7	a_{71}	a_{72}
V8	a_{81}	a_{82}

In altre parole, questa matrice esprime le correlazioni tra le variabili osservate e le variabili latenti che vengono identificate tramite i processi dell'analisi fattoriale.

Quindi il modello dell'analisi fattoriale ha l'obiettivo di spiegare le correlazioni tra le variabili osservate, cioè la varianza comune alle variabili, in funzione di un numero limitato di fattori non osservabili, in modo tale che l'agglomerato delle osservazioni possa essere trasformato in una struttura semplice, ma informativa quasi quanto quella di partenza.

In letteratura la varianza spiegata dalle variabili latenti viene detta **comunanza**; tuttavia va tenuto conto che non tutta la varianza è condivisa tra le variabili: infatti ogni variabile è caratterizzata da una porzione di variabilità che può non venire spiegata dai fattori comuni.

Lo schema seguente è una rappresentazione grafica che mostra che 2 variabili osservate V1 e V2 sono influenzate dai fattori comuni F1 e F2, che spiegano la comunanza delle due variabili in misura proporzionale alle saturazioni a_{ij} (con $i = 1, 2$ e $j = 1, 2$), e dai fattori specifici u1 e u2 relativi alle variabili.



Ricapitolando, i **fattori comuni** (o cioè i fattori latenti) F rappresentano ciò che è in comune a tutte le variabili in analisi e sono detti comuni proprio perché influenzano più di una variabile.

Le relazioni tra le variabili e i fattori vengono definite **saturazioni**. In particolare le saturazioni rappresentano le correlazioni tra variabili e fattori.

L'**unicità** delle variabili, cioè quella parte di varianza specifica di ciascuna variabile è dovuta a cause sistematiche specifiche, o all'errore casuale di misurazione. Essa viene indicata con u_i , dove $i = 1, 2, \dots, p$ (p è il numero di variabili osservate).

Formalmente nell'analisi fattoriale ogni variabile è espressa come combinazione lineare dei fattori responsabili della correlazione con le altre variabili e dell'unicità specifica della variabile stessa:

$$Z_i = F_{1a_{i1}} + F_{2a_{i2}} + \dots + F_{ma_{im}} + u_{i c_i}, \quad i = 1, \dots, p$$

Gli assunti dell'analisi fattoriale sono:

- $\text{Cov}(u_i, F_j) = 0$ cioè l'indipendenza tra i fattori e le unicità;
- $\text{Cov}(u_i, u_j) = 0$ cioè l'indipendenza tra le unicità;
- $\text{Cov}(F_i, F_j) = 0$ che determina l'ortogonalità tra i fattori. Ogni fattore, quindi, spiega caratteristiche del fenomeno in analisi che sono diverse da quelle spiegate da tutti gli altri fattori.

3.2 Comunanza e unicità dei fattori

Dato che le variabili osservate sono funzione di fattori comuni e di un fattore specifico, anche le loro varianze possono essere espresse come composizioni di due parti che corrispondono alla varianza comune e alla varianza specifica di ciascuna variabile.

La varianza totale di una variabile (che standardizzata è uguale a 1) è data dalla somma della varianza spiegata dai fattori comuni (h^2) e la sua unicità u^2 :

$$\text{Varianza totale (standardizzata)} = 1 = h^2 + u^2$$

Grazie agli studi di Thurstone (1931) è possibile ricavare una relazione tra il punto di partenza dell'analisi fattoriale col punto di arrivo, tramite l'uguaglianza

$$R = AA' + U^2,$$

dove:

- A è la matrice delle saturazioni;
- U è la matrice diagonale delle unicità.

Tale relazione spiega la matrice delle correlazioni tra le variabili osservate in funzione delle saturazioni, calcolate durante l'analisi fattoriale, e delle unicità.

La quantità AA' rende ragione degli elementi al di fuori della diagonale di R, mentre U^2 rende ragione degli elementi sulla diagonale, cioè le varianze delle variabili.

3.3 Il processo di estrazione dei fattori

Lo scopo principale dell'analisi fattoriale è quello di estrarre un numero sufficiente di fattori per rendere conto della matrice delle correlazioni tra le variabili osservate R .

Esistono differenti metodi per l'estrazione dei fattori e ognuno di essi è caratterizzato da una procedura che lo contraddistingue.

Infatti ci sono metodi che:

- cercano di spiegare il massimo di varianza comune alle variabili;
- massimizzano la riproduzione di R , cioè vengono estratti fattori che ricostruiscono al meglio l'informazione di partenza;
- richiedono una stima iniziale delle comunanze (varianza in comune);
- richiedono una stima iniziale del numero di fattori da estrarre.

Le saturazioni possono essere ricavate attraverso il calcolo delle *componenti principali*, metodo che però non considera la scissione tra varianza comune e unicità. Le stime delle correlazioni della matrice R vengono quindi gonfiate perché contaminate dalla varianza unica.

Il metodo delle componenti principali si pone come obiettivo quello di costruire delle nuove variabili, ottenute come combinazioni lineari delle variabili originarie, in modo che un numero ridotto di queste nuove variabili sia in grado di spiegare una porzione rilevante della varianza totale.

Ogni componente principale è caratterizzata da una propria varianza λ_i , detta *autovalore* relativo alla componente i -sima ($i = 1, 2, \dots, q$).

La prima componente principale è, tra tutte le componenti principali, quella che spiega più varianza: il suo autovalore è quindi il più grande; la

seconda componente λ_2 , depurata della variabilità spiegata dalla prima, è quella che spiega più varianza possibile dopo la prima e così via.

La varianza totale viene quindi ridistribuita nelle componenti principali, con una forte concentrazione nelle prime componenti: gli autovalori sono più grandi quanto più grande è la porzione di variabilità spiegata. Quindi vale la relazione:

$\lambda_1 > \lambda_2 > \dots > \lambda_r$, dove r è il rango della matrice delle correlazioni.

Inoltre $\sum \lambda_i = \sum \text{var}(v_i)$ che è la varianza totale.

La percentuale di varianza spiegata da una componente principale quindi coincide con il rapporto $\lambda_i / \sum \lambda_i$.

Le componenti principali sintetizzano l'informazione relativa alla varianza delle variabili, e alla correlazione tra le variabili.

Questa operazione non distorce la matrice delle correlazioni R , ma la riassembla.

Solitamente vengono estratte le componenti principali analizzando la matrice delle correlazioni: si cerca cioè di individuare dei fattori che spieghino più varianza possibile delle variabili originarie. Ma invece di analizzare la matrice R completa (ovvero la matrice delle correlazioni con tutti 1 sulla diagonale principale) viene studiata una matrice che sulla diagonale principale contiene delle stime delle comunalità al posto dei valori uguali a 1, eliminando così la componente unica di ogni variabile.

Le stime delle comunalità possono essere fatte in diversi modi, uno di essi consiste nel calcolare, per ogni variabile, la media delle correlazioni con tutte le altre variabili. Questo procedimento si differenzia dall'analisi delle componenti principali classica proprio per il fatto che viene fatta

una distinzione tra varianza comune e varianza specifica. In questo modo la soluzione non viene contaminata dalla varianza unica.

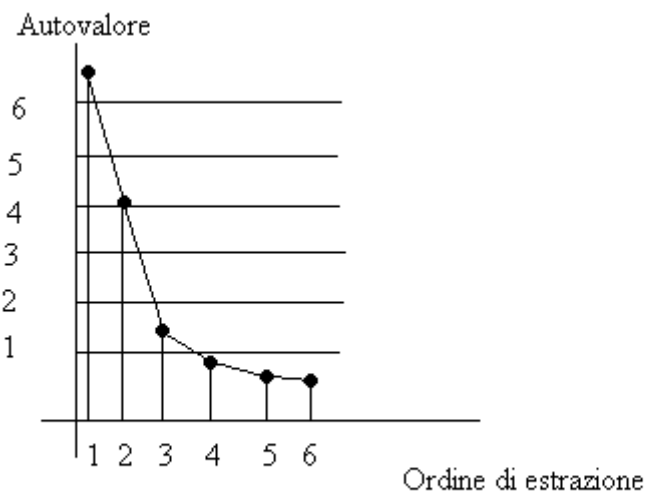
3.4 Scelta del numero di fattori da estrarre

Esistono diversi metodi per decidere quanti fattori estrarre: uno tra i più usati è il metodo basato sulla **rappresentazione grafica degli autovalori** delle componenti principali, detto **metodo del punto di gomito**.

E' possibile rappresentare come punti del piano cartesiano gli autovalori facendo riferimento all'ordine di estrazione. Come spiegato nel paragrafo 3.3, il valore di questi ultimi è grande per i fattori estratti per primi e diminuisce a mano a mano che vengono estratte le componenti principali successive.

Unendo i punti così identificati si ottiene una spezzata: se l'analisi fattoriale è un metodo adatto per l'analisi dei dati, i punti che rappresentano i primi autovalori saranno notevolmente più grandi e distanziati rispetto ai successivi; questo fa sì che i segmenti che li uniscono siano molto inclinati. Gli autovalori che seguono assumono valori decrescenti molto più bassi: sono uniti da segmenti quasi paralleli all'asse delle ascisse, come si vede nell'esempio della figura che segue. Si continua il processo di estrazione dei fattori finché si raggiunge il punto in cui la curva degli autovalori decrescenti cambia pendenza, dando forma a quello che viene definito un "punto di gomito", e diventa sostanzialmente piatta.

Come è spiegato in Fabbris (1990) solitamente si usa considerare solo i fattori il cui autovalore è maggiore di 1.



Come si vedrà nel paragrafo 3.6, l'applicazione di questo criterio, che non si basa sulla percentuale di varianza spiegata, consente estrarre un esiguo numero di fattori.

Un altro criterio per decidere il numero di fattori da estrarre è quello della **varianza spiegata dai fattori** che consiste nell'estrazione di un numero di fattori sufficientemente elevato da permettere di spiegare una percentuale di varianza predefinita. Una quota di varianza spiegata pari al 75% è considerata un buon traguardo; tuttavia spesso sono accettate anche percentuali inferiori a questo valore. Questo metodo garantisce un'analisi accurata (infatti più sono i fattori estratti, più è precisa l'analisi fattoriale), ma può non essere adeguata all'analisi fattoriale da svolgere proprio perché non favorisce l'estrazione di un numero di fattori basso. Ad esempio, nel caso specifico dei dati oggetto di studio in questa tesi, lo scopo è quello di semplificare il fenomeno che si sta studiando, con una riduzione del numero delle variabili consistente (si punta a ricavare 2 o massimo 3 fattori): dal momento che si è interessati a svolgere un'analisi esplorativa può essere preferibile dare precedenza alla semplicità a discapito della precisione.

3.5 I metodi di rotazione dei fattori

Una volta calcolati i pesi fattoriali è necessario capire quali sono gli aspetti effettivamente descritti dai fattori per comprendere le caratteristiche del fenomeno che è stato sintetizzato.

Con questo scopo si attua la **rotazione dei fattori** che è necessaria per semplificare l'interpretazione della soluzione.

La rotazione consiste nel modificare la matrice delle saturazioni e si basa sul fatto che esistono infinite matrici A che permettono di riprodurre altrettanto bene la matrice R delle correlazioni. Infatti la soluzione ruotata e quella non ruotata spiegano la stessa porzione di varianza.

Esistono 2 tipi di rotazione:

- le *Rotazioni ortogonali* che rispettano gli assunti dell'analisi delle correlazioni; in particolare obbediscono alla condizione di ortogonalità tra gli assi fattoriali;
- le *Rotazioni oblique* che ruotano gli assi permettendo che i fattori siano tra loro correlati. Non rispettano quindi l'assunto di ortogonalità.

Le rotazioni sono basate sulla logica della *struttura semplice di Thurstone*, che fa in modo che le variabili cadano il più vicino possibile agli assi fattoriali massimizzando il numero di saturazioni prossime allo 0 per ogni fattore. Questo è attuabile perché ciascun fattore è correlato con un numero limitato di variabili.

I principali criteri di rotazione dei fattori sono Varimax, Quartimax, Oblimin e Promax.

Varimax

E' un tipo di rotazione ortogonale che aumenta la semplicità dei fattori massimizzando la varianza delle saturazioni delle variabili all'interno di ogni fattore.

Per ogni fattore tende a far aumentare le saturazioni già elevate e a far diminuire quelle più basse. In questo modo si avranno solo saturazioni prossime a 1 e saturazioni prossime a 0.

Quartimax

E' un altro tipo di rotazione ortogonale che massimizza la varianza di ogni saturazione. Tende a facilitare la formazione di un unico fattore generale.

Oblimin

E' un metodo di rotazione obliqua. Fa in modo che le variabili abbiano saturazioni vicine a 0 in tutti i fattori tranne in uno. Considera sia le varianze che le covarianze tra i fattori.

Promax

E' un tipo di rotazione obliqua che partendo da una rotazione ortogonale Varimax la modifica per renderla più semplice aumentando i pesi fattoriali già grandi e riducendo quelli più piccoli, consentendo che i fattori siano correlati. Fornisce risultati simili a Oblimin.

3.6 Analisi dei dati

L'analisi fattoriale è un'applicazione che permette di esplorare i dati in base alle relazioni che intercorrono tra le variabili che sono state osservate.

Nel caso dei dati sulla percezione del burnout dei docenti viene adottata la tecnica dell'analisi fattoriale per esaminare se vi sono classi di variabili riassumibili con dei fattori unici, ricavati col metodo delle componenti principali (spiegato nel paragrafo 3.3), che mostrino se vi sono comportamenti comuni a più unità statistiche e che verifichino l'esistenza di gruppi di unità differenti.

Si ricorda che uno degli scopi principali dell'analisi sui dati del burnout è la ricerca delle differenze tra la percezione del burnout nelle scuole pubbliche e quelle private.

Dalle analisi descrittive preliminari è apparso che i sintomi del burnout, per quanto poco frequenti e non allarmanti, sono percepiti meno raramente nel sottocampione della scuola pubblica che non in quello della scuola privata.

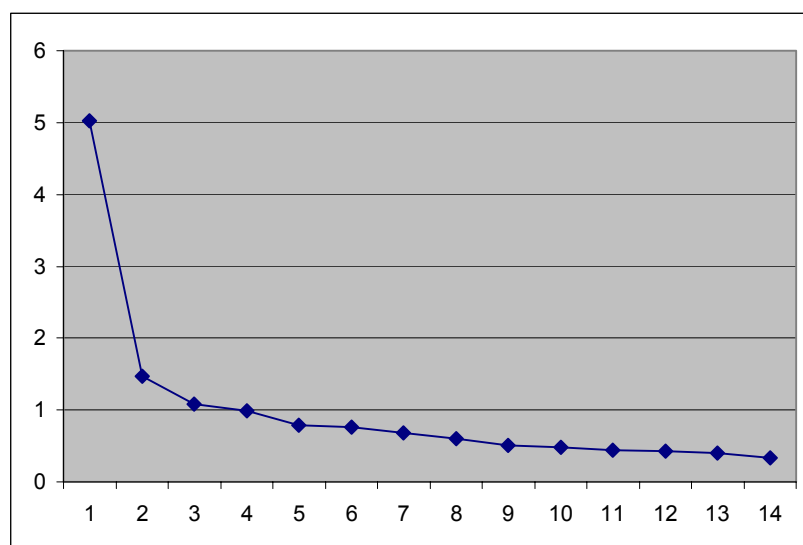
Quindi lo scopo dell'analisi fattoriale esplorativa sarà quello di constatare se vengono riscontrate le affermazioni fatte nel secondo capitolo.

Le 14 variabili che compongono il questionario sono state trattate come variabili numeriche discrete che possono assumere valori da 1 a 5. La matrice R delle correlazioni tra queste variabili è il punto di partenza per l'analisi fattoriale.

Le restanti 2 variabili categoriali che descrivono il sesso e il tipo di scuola del rispondente non possono essere incluse nel calcolo dei fattori comuni; sono state usate come variabili passive che fungono da etichetta per i punteggi fattoriali delle unità statistiche.

Una volta tracciato il diagramma dei punteggi fattoriali sul piano dei primi due assi fattoriali, tali etichette permettono di identificare possibili analogie tra la posizione rispetto agli assi e il tipo di scuola o il sesso.

Il grafico che segue rappresenta la spezzata degli autovalori.

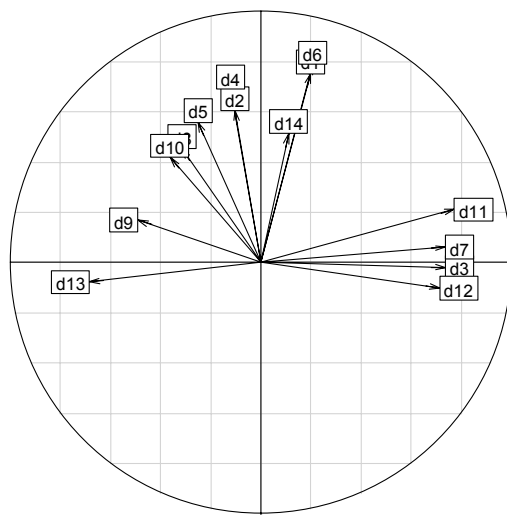


Si nota che il primo e il secondo autovalore sono gli unici che possono essere considerati rilevanti dato che sono chiaramente più elevati dei successivi. Dal terzo in poi infatti la spezzata diventa quasi piatta. La presenza di fattori che prevalgono sugli altri, con autovalori nettamente superiori a 1, conferma che l'analisi fattoriale è un metodo adeguato per analizzare i dati in questione.

Quindi nell'analisi fattoriale che si vuole effettuare, l'obiettivo è spiegare più varianza comune alle variabili possibile, con due sole variabili latenti.

Definito il numero delle dimensioni col comando `dudi.pca()` (si veda l'Appendice del Capitolo 3 per una guida approfondita su come è stato usato il software R) si ricavano i pesi fattoriali e i punteggi fattoriali.

Ruotando i fattori con il metodo Promax è possibile ottenere dei pesi fattoriali delle variabili che favoriscono una semplice interpretazione dei 2 fattori: infatti, osservando il cerchio delle correlazioni che segue, si nota che un gruppo di variabili correla quasi unicamente col primo fattore (d3, d7, d11, d12, d9, d13); un altro gruppo invece correla quasi unicamente col secondo fattore (d1, d2, d4, d5, d6, d14).



Variabile	Affermazione corrispondente
d1	Spossatezza dopo il lavoro
d2	Esaurito prima del lavoro
d3	Molto soddisfatto
d4	Sfinimento dovuto agli allievi
d5	Frustrato
d6	Insegnare è pesante
d7	Sfrutta appieno le sue capacità
d8	Molto teso
d9	Considera di abbandonare l'insegnamento
d10	Impaziente
d11	Sceglierebbe di nuovo l'insegnamento
d12	Si sta realizzando nella vita
d13	Sprecato per l'insegnamento
d14	Non è qualificato per insegnare

Esistono delle linee guida che permettono di definire la rilevanza della correlazione tra un fattore e una variabile (cioè la saturazione) e quindi di capire quali sono le variabili influenzate maggiormente da tale fattore. Tale criterio si basa sulla dimensione della saturazione a_{ij} tra fattore comune j -simo e variabile osservata i -sima, che, elevata al quadrato, corrisponde alla quota di varianza della variabile che il fattore riesce a spiegare.

La tabella che segue spiega come viene definita la qualità delle saturazioni in funzione del modulo della loro grandezza.

Così, ad esempio, se un peso fattoriale (o saturazione) a_{ij} è uguale a 0.60 viene considerato buono perché il fattore j -simo spiega almeno il 30% della varianza della variabile i -sima, che è una quota abbastanza grande: questo equivale a dire che il fattore j -simo riproduce discretamente l'informazione fornita della variabile i -sima originale.

Dunque, più grande è il valore assoluto della saturazione a_{ij} ($|a_{ij}|$), più è determinante la variabile osservata i -sima nella costruzione del fattore j -simo e viceversa.

Grandezza saturazione	Percentuale di varianza spiegata	Qualità
$ a_{ij} > 0.71$	50%	Eccellente
$0.63 < a_{ij} < 0.71$	40%	Molto buona
$0.55 < a_{ij} < 0.63$	30%	Buona
$0.45 < a_{ij} < 0.55$	20%	Sufficiente
$0.32 < a_{ij} < 0.45$	10%	Scarsa
$ a_{ij} < 0.32$	< 10%	Inadeguata: può essere ignorata

La tabella che segue mostra i pesi fattoriali nel modello di analisi fattoriale ricavato, ottenuti con rotazione Promax.

I due fattori sono indicati con F1 e F2 e le 14 variabili con D1, D2 ..., D14)

	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14
F1	0.19	-	0.73	-	-		0.73	-	-	-	0.77	0.71	-	0.11
F2	0.75	0.10		0.12	0.25	0.20		0.31	0.48	0.36			0.68	
		0.60		0.69	0.55	0.78		0.45	0.16	0.41	0.21	-		0.51
												0.10		

Sulla base di quanto spiegato dalla tabella delle linee guida sono state evidenziate le saturazioni eccellenti e molto buone. Così si vede che il

primo fattore sintetizza in modo eccellente le variabili d3, d7, d11, d12 (dato che le relative saturazioni sono almeno pari a 0.71) che corrispondono alle 4 domande del questionario che misurano la soddisfazione che il docente di educazione fisica riceve grazie al suo lavoro. Inoltre il primo fattore sintetizza in maniera molto buona di varianza della variabile D13 (dato che $|-0.68| = 0.68$ è più grande di 0.63).

L'interpretazione dei fattori viene effettuata andando a controllare cosa esprimono le variabili che correlano con essi in modo considerevole.

Il primo fattore comune può quindi essere interpretato come misura della soddisfazione e del gradimento del lavoro che i docenti manifestano agli alunni.

Si osserva dal cerchio delle correlazioni che la variabile d13 è correlata negativamente con il primo fattore e la direzione del relativo vettore è quasi coincidente con quella del vettore d7, ma con verso opposto. In particolare, leggendo le domande del questionario riferite a queste variabili ci si accorge che la domanda 13 “Pensa che in un'altra professione (diversa dall'insegnamento) potrebbe sfruttare in modo migliore le sue capacità intellettive” esprime un concetto quasi contrario a quello espresso nella domanda 7 “Sente che l'insegnamento le/gli (al docente) permette di sfruttare al meglio le sue capacità”. Il confronto di come correlano queste due variabili con il fattore fa capire che l'informazione che forniscono le due variabili, nell'ambito dell'analisi fattoriale che sto svolgendo, è praticamente la stessa: le due variabili, cioè, forniscono informazioni opposte.

Il secondo fattore influenza in modo consistente le variabili d1, d4 e d6; inoltre spiega una dose discreta di varianza delle variabili d2 e d5. Queste variabili fanno tutte riferimento alla fatica legata all'insegnamento: il secondo fattore può quindi essere interpretato come una misura dello

sforzo e dello stress che gli scolari percepiscono nel professore di educazione fisica.

Una ulteriore considerazione riguarda l'incorrelazione tra le variabili che determinano fattori diversi. Osservando il cerchio delle correlazioni, infatti, si nota che le variabili influenzate dal primo fattore e quelle influenzate dal secondo, sono tra loro ortogonali. Questo significa che i fenomeni descritti dai due gruppi di variabili sono indipendenti, cioè il modo di manifestarsi delle variabili del primo gruppo non influisce sul modo di manifestarsi di quelle del secondo gruppo.

L'assenza di correlazione tra queste variabili è osservabile anche dalla matrice delle correlazioni R.

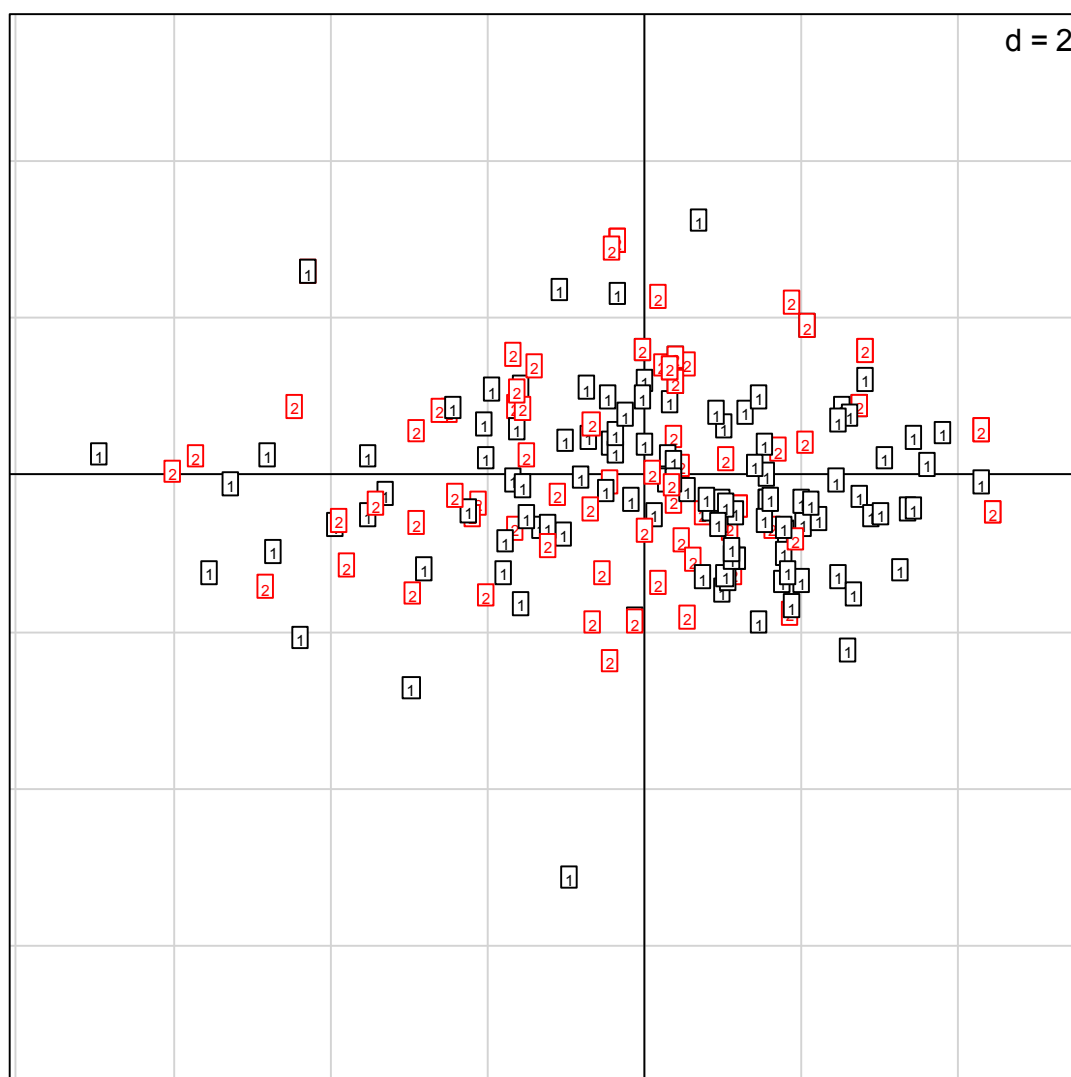
In base all'interpretazione dei fattori che è stata eseguita l'incorrelazione tra le variabili che influenzano il primo e il secondo fattore ci dice che non vi è un legame tra lo stress dei professori e il gradimento per il lavoro che svolgono.

In altre parole non è detto che un professore affaticato a causa del proprio lavoro, al punto di ritenere frustrante il contatto con gli alunni, provi il desiderio di cambiare professione perché non è soddisfatto dalla sua attuale posizione; allo stesso modo non si può affermare che un professore che ama il proprio lavoro e che trae soddisfazione da esso, non sia snervato.

La percentuale di varianza spiegata dai due fattori è pari a 45,6%, con i risultati della rappresentazione grafica dei pesi e dei punteggi fattoriali che sembra essere buona e carica di significato.

Come ultimo commento, è importante sottolineare che è stata eseguita un'analisi di sensibilità dell'analisi fattoriale a partire da un dataset in cui i valori delle modalità di risposta sono stati riscritti (cioè le variabili non assumono più valori 1, 2, 3, 4 e 5 ma ad esempio 10, 20, 30, 40, 50): in

questo modo varia la scala di misura, ma le distanze tra le modalità sono proporzionali a quelle originali. Il risultato è stato uguale a quello dell'analisi fattoriale sui dati originali. Questo significa che la trasformazione delle variabili ordinali in variabili numeriche non ha causato distorsione nei risultati (cfr. Agresti, Finlay (1997))



Dal grafico dei punteggi fattoriali si nota, come era stato preventivato nelle analisi descrittive, che non vi è una netta separazione delle unità

della scuola privata (unità rosse etichettate con 2) e quelle della scuola pubblica (unità nere etichettate con 1).

Infatti le unità sono sparpagliate in modo abbastanza omogeneo intorno all'origine del sistema di assi fattoriali.

Tuttavia è presente una concentrazione più accentuata di punteggi fattoriali in prossimità dell'origine e nel quadrante in basso a destra. Questo è un buon segno perché significa che all'interno del campione è diffusa l'idea che i professori sono abbastanza soddisfatti del loro lavoro e che sopportano bene la fatica che esso comporta. In altre parole diversi alunni ritengono che il proprio professore di educazione fisica apprezzi il proprio lavoro e lo svolga volentieri.

La tendenza dei punteggi fattoriali di essere collocati in prossimità dell'origine degli assi può essere interpretata come segnale che generalmente i professori oggetto dell'indagine non manifestino comportamenti anomali che facciano trasparire evidenti tracce di stress tipiche della sindrome di burnout, ma allo stesso tempo non mostrano di amare e svolgere tanto volentieri il proprio lavoro.

L'analisi fattoriale che è stata svolta non coglie le differenze tra scuola pubblica e scuola privata che erano state riscontrate nell'analisi descrittiva.

3.7 Analisi cluster applicata al modello stimato con l'analisi fattoriale

Dopo aver eseguito l'analisi fattoriale può essere d'aiuto operare un'analisi cluster per classificare le unità sulla base dei valori assunti dai fattori comuni. Lo scopo di questa procedura è formare dei gruppi "omogenei" secondo un certo criterio, a cui vengono attribuite un certo numero di caratteristiche attinenti a tutti i componenti del gruppo, e solo di questi (almeno una caratteristica deve differire da gruppo a gruppo).

I gruppi o cluster sono a priori incogniti, sia di numero (cioè non sappiamo prima di iniziare l'analisi quanti gruppi saranno necessari per una "buona" classificazione), sia di "posizione", nel senso che non sappiamo quali siano le caratteristiche che li differenziano.

Per un più approfondito studio della cluster analysis si rimanda a Fabbris (1990).

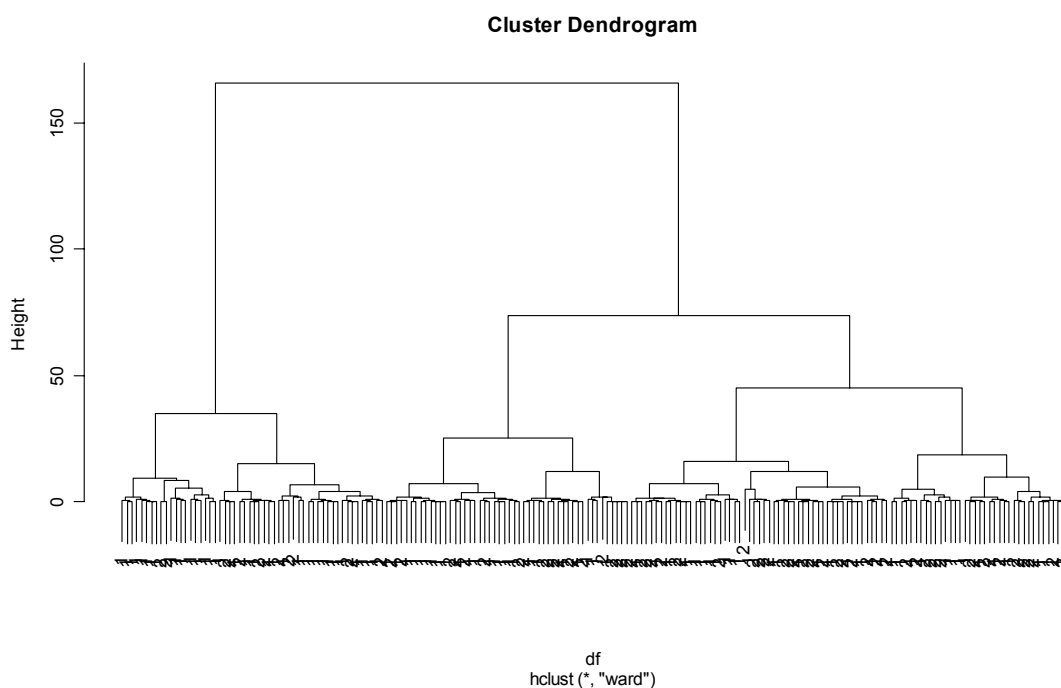
Tramite questa tecnica si può cercare di capire se i punteggi fattoriali calcolati riescono a cogliere le differenze tra gli eventuali sottogruppi di unità (nel nostro caso alunni della scuola privata e alunni della scuola pubblica), evidenziate dalle analisi preliminari e dalla precedente analisi cluster.

L'analisi della matrice delle variabili latenti viene eseguita con un **metodo gerarchico agglomerativo**, che consente di effettuare un'operazione di raggruppamento delle unità campionarie che hanno dato risposte più simili alle domande del questionario. Un metodo gerarchico aggregativo consiste nel ridurre progressivamente il numero dei cluster unendo quelli più simili in base a una misura di "vicinanza" (o somiglianza), a partire da un numero di gruppi pari al numero di dati (quindi ogni cluster iniziale contiene un solo dato).

La procedura di raggruppamento dei dati può essere fatta con metodi diversi. Nel caso in questione è stato adottato il metodo di *Ward*, che prevede che i gruppi vengano aggregati in modo che l'incremento di varianza nei nuovi gruppi sia più piccola possibile (tutte le volte che si aggiunge un'unità ad un gruppo la varianza aumenta: il criterio di Ward minimizza questo aumento unendo i gruppi più simili).

Dato che i punteggi fattoriali possono essere considerati variabili numeriche le differenze tra le unità sono distanze euclidee.

Il dendrogramma che risulta è il seguente.



Il numero ottimo di gruppi che vanno considerati può essere determinato ispezionando il dendrogramma. Quando si effettua un'analisi gerarchica infatti il numero di gruppi può essere deciso sezionando il dendrogramma all'altezza del massimo salto tra livelli di somiglianza ai quali sono

avvenute le aggregazioni dei cluster. Le altezze (che nel grafico sono riportate sulla retta verticale a sinistra del dendrogramma) misurano la consistenza della dissomiglianza tra i gruppi. Quindi tanto maggiore è la differenza di altezza tra i gruppi e tanto più sono diversi tra loro.

Nel caso in analisi il massimo salto tra livelli di somiglianza ai quali sono avvenute le aggregazioni dei cluster corrisponde alla differenza tra 130 e 75, cioè tra le altezze dei due nodi più alti nel grafico.

In questo modo si ottengono 2 cluster che però non sono omogenei al loro interno rispetto all'attributo "tipo di scuola". Questa particolarità è una conferma di quanto affermato nei commenti sull'analisi fattoriale, riguardo alla capacità di raggruppare gli alunni in base ai fattori. Infatti anche questa tecnica di esplorazione dei dati mostra che i fattori non distinguono in modo soddisfacente le unità di scuole diverse.

Si nota che uno dei due gruppi è molto più numeroso dell'altro e in particolare il gruppo meno numeroso (le cui unità nel grafico sono quelle a sinistra) è composto prevalentemente da unità che si riferiscono ad alunni della scuola pubblica. Solo nella formazione di questo gruppo l'algoritmo di clusterizzazione ha colto alcune differenze tra i due sottocampioni che i fattori rilevano, seppure non chiaramente.

Tuttavia se il dendrogramma viene analizzato capillarmente, osservando i cluster alle altezze più basse, si nota che le unità sono suddivise in molti gruppi, composti da poche unità, che sono omogenei per il carattere "tipo di scuola".

Dunque si ipotizza che le variabili del questionario possano effettivamente fornire dei criteri che permettano di suddividere alunni della scuola privata e alunni della scuola pubblica e che la sintesi dell'analisi fattoriale non riesce a cogliere.

Un'analisi cluster applicata alle variabili osservate e non ai fattori può verificare tale ipotesi.

3.8 Analisi cluster applicata ai dati originari

Un'analisi cluster applicata ai dati può dare prova del fatto che vi siano differenze tra la percezione della sindrome di burnout nelle scuole pubbliche e in quelle private.

Nel caso oggetto di studio si considerano soltanto le variabili d3, d5, d7, d8, d9, d10, d12, d14 sulla base delle quali verranno determinate le differenze (o distanze) tra le unità statistiche. Gli allievi (o unità) che hanno risposto in modo simile alle domande verranno quindi raggruppati in clusters: quindi si osserverà se i gruppi formati sono composti prevalentemente da alunni della scuola privata o della scuola pubblica. In altre parole lo scopo è quello di osservare se gli alunni del primo e del secondo tipo di scuola si differenziano nel modo percepire lo stato di salute mentale del proprio professore di educazione fisica. Si è scelto di considerare solo le variabili che nell'analisi preliminare (vedi cap2) hanno manifestato le differenze tra i due sottocampioni di allievi delle scuole pubbliche e delle scuole private.

Questa decisione è stata presa in modo da semplificare l'analisi cluster: infatti sono escluse le informazioni non utili (cioè quelle fornite dalle restanti variabili) per captare le differenze tra unità della scuola privata e unità della scuola pubblica.

Va sottolineato che questa analisi cluster si distingue da quella precedente per il fatto che viene considerato un gruppo di variabili diverso: in questo caso infatti non sono state adoperate le informazioni provenienti

da tutte le variabili del questionario (come invece è successo nell'analisi cluster applicata alla matrice dei punteggi fattoriali). Questa diversità può determinare risultati molto diversi tra le due analisi cluster, anche se il fenomeno che si vuol analizzare è lo stesso.

Inoltre non si è interessati ad identificare quali siano le variabili che classificano e quindi a capire quali sono le caratteristiche di ciascun gruppo che verrà a formarsi: queste informazioni infatti sono già state ricavate dalle analisi preliminari e dall'analisi fattoriale. Nel caso in questione l'analisi cluster viene usata come uno strumento che permetta di assegnare le unità a classi multivariate non predefinite: in questo modo, dopo la formazione dei gruppi, si potrà osservare se ogni cluster è omogeneo al suo interno rispetto alla variabile "tipo di scuola". Infatti ci si aspetta che vengano composti gruppi in cui prevalgono gli allievi della scuola privata, e gruppi in cui prevalgono gli allievi della scuola pubblica, in modo da rendere evidente l'esistenza di differenze tra i due sottocampioni.

E' stato scelto di utilizzare una tecnica gerarchica agglomerativa perché è un tipo di analisi piuttosto rapido da svolgere rispetto alla maggior parte delle altre tecniche di clusterizzazione.

Generalmente per ottenere una classificazione con gruppi molto omogenei, vengono scelte le tecniche di classificazione **non gerarchiche**, che prevedono che sia noto a priori il numero di gruppi K che si vogliono ottenere, e poi che si "lasci andare" l'algoritmo di classificazione, che presenterà come output finale la divisione dei dati in K cluster. Tuttavia, come è spiegato in Fabbris (1990), anche il metodo gerarchico aggregativo del legame completo è un metodo idoneo all'individuazione di gruppi omogenei. Ci si aspetta quindi che i risultati ottenuti applicando tale metodo ai dati analizzati in questa tesi siano soddisfacenti.

Inoltre le analisi cluster non gerarchiche sono applicate preferibilmente nei casi in cui si hanno abbastanza informazioni sui dati da poter fare delle ipotesi sul numero di gruppi di unità omogenei che possono esistere nel campione in analisi.

Nel caso dei dati sulla sindrome di burnout che si stanno analizzando non possiamo fare una stima del numero dei gruppi. Pur avendo dimostrato che ci sono delle differenze tra i due eventuali gruppi della scuola pubblica e della scuola privata, dall'analisi fattoriale è emerso i sottocampioni dell'uno e dell'altro tipo di scuola sono composti da unità eterogenee. E' quindi ipotizzabile che esistano ulteriori sottogruppi di unità omogenee all'interno dei due sottocampioni dei quali finora non abbiamo saputo definire l'entità.

Per applicare una tecnica non gerarchica è bene che sugli stessi dati siano state condotte altre analisi, ad esempio una classificazione di tipo gerarchico, per identificare dei gruppi che fungano da soluzione di partenza.

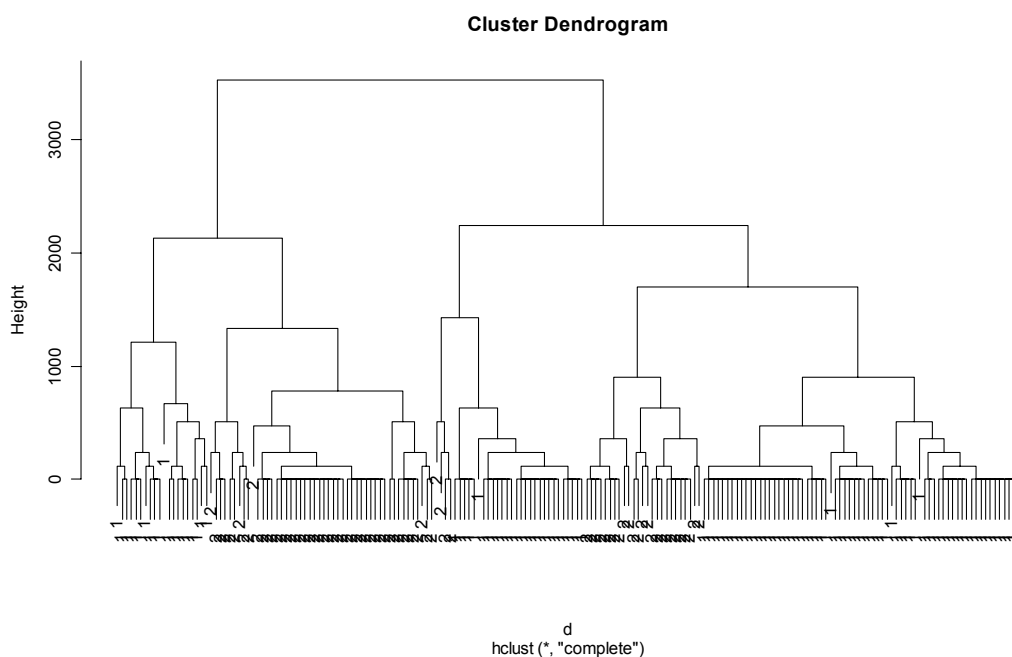
Le distanze (cioè le differenze) tra le unità sono state calcolate col metodo della *distanza di media assoluta* che è un tipo di distanza particolarmente adatta nel caso in cui le variabili che caratterizzano le osservazioni siano su scala ordinale.

Essa viene comunemente detta *distanza di manhattan* d_{hk} e viene calcolata, tra le unità h e k nello spazio p -dimensionale definito dalle p variabili osservate, con la formula:

$$d_{hk} = \sum |x_{hi} - x_{ki}|, \quad h \neq k = 1, 2, \dots, n \quad (\text{dove } n \text{ è il numero delle osservazioni})$$
$$i = 1, \dots, p \quad (\text{dove } p \text{ è il numero delle variabili considerate})$$

Il risultato dell'analisi cluster gerarchica aggregativa che ne scaturisce è rappresentata dal dendrogramma che segue.

Come nell'analisi fattoriale, le unità sono state etichettate con 1 o 2 a seconda del tipo di scuola cui appartengono (si ricorda che 1 corrisponde alla scuola pubblica e 2 alla scuola privata).



Nel caso in questione è evidente che la massima differenza di altezza tra nodi è approssimativamente tra 2100 e 3500, nella parte più alta del dendrogramma. In base a questo criterio vengono considerati 2 soli gruppi che si formano all'altezza 3500. All'interno del gruppo che nel grafico si trova a sinistra, prevalgono per numero gli alunni della scuola privata, mentre nel gruppo di destra prevalgono gli alunni della scuola pubblica. E' evidente che esistono degli ulteriori sottogruppi di unità che si differenziano in base alla variabile "tipo di scuola". Infatti aumentando il numero dei cluster per un'analisi più profonda che permetta di cogliere più categorie di dati, è possibile identificare 5 gruppi perfettamente o-

moganei in base al tipo di scuola, visibili in modo chiaro anche nel dendrogramma, che si formano in prossimità delle altezze 2100 e 1600.

Come ulteriore commento, va osservato che con questa analisi cluster sono stati identificati dei gruppi di alunni appartenenti allo stesso tipo di scuola che non si era stati in grado di osservare tramite l'analisi fattoriale.

La netta suddivisione delle unità in gruppi di allievi omogenei secondo il tipo di scuola frequentata è prova che le variabili utilizzate per la classificazione colgono bene le differenze tra l'intensità con cui i professori manifestano lo stress o l'appagamento dovuto al proprio lavoro.

La separazione degli alunni appartenenti allo stesso tipo di scuola in gruppi diversi può essere dovuta al fatto che i professori a cui gli alunni hanno fatto riferimento per compilare il questionario sono persone diverse, con stati di salute mentale diversi.

APPENDICE DEL CAPITOLO 3

Comandi per l'analisi fattoriale

In questa sezione vengono riportati i comandi del software R utilizzati per svolgere l'analisi fattoriale.

Inizialmente è necessario caricare la libreria specifica che contiene le funzioni per svolgere l'analisi fattoriale (ade4).

```
library(ade4)
```

comando che carica la libreria ade4 nella sessione di lavoro in corso.

Per maggiori informazioni sulla libreria ade4 visionare

<http://cran.r-project.org/doc/packages/ade4.pdf>

Il primo passo dell'analisi fattoriale è la creazione del modello sulla base del dataset che dobbiamo analizzare (nel caso specifico il dataset è stato chiamato “dati”).

Questa operazione avviene col comando

```
fit=dudi.pca(dati[,4:17])
```

che crea un modello, denominato “fit”, di analisi fattoriale ricavando le componenti principali (fattori comuni) sulla base delle 14 variabili del questionario. `dati[,4:17]` indica che vengono considerate tutte le osservazioni del campione e le variabili dalla 4° alla 17° (che sono appunto quelle del questionario) dal dataset “dati”.

Immediatamente R mostra il grafico degli autovalori di tutte le componenti principali e richiede di specificare il numero di fattori da estrarre (che viene scelto con uno dei metodi descritti nel paragrafo 3.4).

Nel caso specifico dei dati sulla percezione del burnout sono stati estratti 2 fattori.

Successivamente alla creazione dei due fattori per opera del comando `dudi.pca()` è necessario ricercare una loro interpretazione ragionevole studiando il significato delle variabili che correlano maggiormente con essi.

A questo scopo viene osservato il cerchio delle correlazioni che si crea col comando

```
s.corcircle(fit$co)
```

che disegna il cerchio delle correlazioni tra le variabili e i fattori ricavati precedentemente. `fit$co` è la matrice delle saturazioni.

Spesso è necessario effettuare una rotazione dei fattori che ri assembli la matrice delle saturazioni per facilitare l'interpretazione delle variabili latenti. Si ricorda che questa operazione non ha nessuna ripercussione sulla precisione del modello di analisi fattoriale precedentemente ricavato.

La rotazione promax, che è quella usata nel caso specifico, viene applicata col comando

```
rot=promax(as.matrix(fit$co))
```

che crea un item chiamato "rot". L'oggetto della rotazione `promax()` è la matrice delle saturazioni di "fit". La scrittura `as.matrix(fit$co)` è necessaria per far capire al software che deve considerare come una matrice l'oggetto `fit$co`.

Può quindi venir ridisegnato il cerchio delle correlazioni sulla base degli assi fattoriali ruotati con il comando

```
s.corcircle(rot$loadings[,1:2])
```

che mostra più chiaramente i legami tra variabili e fattori comuni.

`rot$loadings[,1:2]` è la matrice dei nuovi pesi fattoriali che viene rappresentata nel cerchio sotto forma di vettori.

Ora non rimane che rappresentare graficamente i punteggi fattoriali sul piano identificato dai due assi fattoriali col comando

```
s.class(fit$li, fac=as.factor(rownames(dati)), clabel=.5, label=as.factor(dati$SC)).
```

Dal punto di vista pratico questo comando usa i punteggi fattoriali di ogni unità statistica, relativi ai primi 2 fattori comuni, come se fossero le coordinate di un punto raffigurabile sul piano cartesiano. Quindi, una volta tracciati come riferimento i due assi fattoriali, le unità statistiche vengono raffigurate sotto forma di punti, in base all'intensità con cui manifestano ciascuno dei due fattori comuni.

`fit$li` sono i punteggi fattoriali di tutte le unità; `clabel` definisce la dimensione dei rettangoli numerati che rappresentano ciascuna unità; `label` serve per dare un'etichetta alle unità: nel nostro caso ad esempio, ci permette di classificare con un 1 le unità che appartengono alla scuola pubblica e con 2 le unità che appartengono alla scuola private. L'uso di queste etichette permette di osservare se vi sono differenze tra unità di scuole.

Comandi per l'analisi cluster

In questa sezione vengono riportati i comandi per svolgere l'analisi cluster.

Per svolgere l'analisi cluster è necessario caricare nel workspace di R la libreria *cluster*, col comando

```
library(cluster)
```

Dato che l'analisi cluster viene svolta a partire dalle distanze tra le unità va creato un item, che chiamiamo *d*, composto dalle distanze tra le coppie di unità del campione.

Questo avviene col comando

```
d = dist(dati[,c(d3, d5, d7, d8, d9, d10, d12, d14)], method = 'manhattan')
```

che calcola le *distanze di manhattan* tra le unità, in funzione delle variabili d3, d5, d7, d8, d9, d10, d12, d14.

Ora non rimane che applicare l'algoritmo dell'analisi gerarchica aggregativa che raggruppa le unità in clusters e dare un nome all'output di R (in questo caso viene assegnato il nome *clu*) col comando

```
clu = hclust(d, method = 'complete')
```

che raggruppa le unità sulla base delle distanze *d* col metodo del legame completo (come specificato da `method = 'complete'`).

Per rappresentare il dendrogramma (cioè l'albero di classificazione) si utilizza il comando

```
plot(clu, label = SC)
```

dove `label = SC` è il comando che assegna alle unità l'etichetta che specifica il tipo di scuola frequentato (privata o pubblica)

CONCLUSIONI

Dalle analisi fatte è emerso che i professori che sono stati oggetto dell'indagine hanno manifestato di non essere affetti dalla sindrome di burnout. Questo si è visto sia dalle analisi preliminari che dall'analisi fattoriale che hanno segnalato che in generale gli alunni percepiscono che i professori sono abbastanza soddisfatti del loro lavoro e non sono stressati.

Grazie alla sintesi delle informazioni ricavata con l'analisi fattoriale, è apparso che vi sono delle variabili, relative ad alcune domande del questionario, che sono molto più importanti di altre per definire l'intensità con cui gli allievi percepiscono lo stato di salute mentale del proprio professore di educazione fisica. Queste sono: "1) Si sente "a terra" alla fine di una giornata intera a scuola"; "3) Sente che insegnare le/gli dà una grande soddisfazione"; "4) Sente che lavorare con gli studenti per un giorno intero è uno sforzo opprimente"; "6) Sente che lei/ lui deve lavorare troppo duramente nell'insegnamento"; "7) Sente che l'insegnamento le/ gli permette di utilizzare al massimo le sue capacità"; "11) Pensa che se potesse ricominciare la sua vita sceglierebbe di nuovo l'insegnamento"; "13) Pensa che in un'altra professione (diversa dall'insegnamento) potrebbe meglio utilizzare le sue capacità intellettive".

Volendo semplificare la struttura del questionario, è possibile ridurre il numero di domande senza perdere informazioni sull'intensità con cui viene percepita la sindrome di burnout, formulando un questionario composto solamente dalle domande sopraelencate e dal tipo di scuola.

Per quanto riguarda lo studio delle differenze tra i sottocampioni della scuola privata e della scuola pubblica, dall'analisi preliminare è emerso

che, pur essendo evidente che in entrambi i casi i professori oggetto di studio non sono affetti da sintomi della sindrome di burnout, i professori delle scuole private hanno manifestato di ricevere più soddisfazione dal proprio lavoro che non i colleghi delle scuole pubbliche; inoltre gli allievi delle scuole private hanno dimostrato di percepire meno intensamente i sintomi di stress del professore rispetto agli alunni delle scuole pubbliche.

Applicando l'analisi cluster ai dati, si è stati in grado di identificare dei gruppi di unità, dividendo alunni della scuola pubblica e alunni della scuola privata. Questo è stato possibile poiché, per determinare le distanze tra le unità, si sono utilizzate le variabili d3, d5, d7, d8, d9, d10, d12, d14, che, nelle analisi preliminari, hanno dato prova dell'esistenza di diversità tra i sottocampioni della scuola pubblica e della scuola privata.

Quanto affermato nelle analisi preliminare sulle diversità tra professori di scuole diverse è stato quindi confermato anche dall'analisi cluster.

I risultati ottenuti dallo studio sono soddifacenti viste le informazioni che sono state ricavate.

Nell'Introduzione era stato affermato che si voleva verificare l'ipotesi che i casi di burnout sono più frequenti nelle scuole private che non nelle scuole statali, come è stato suggerito da riferimenti bibliografici che riportano testimonianze di studi internazionali sulla sindrome di burnout nei professori delle scuole.

Alla luce delle analisi svolte, invece, verrebbe da affermare il contrario.

Le conclusioni a cui si è giunti sono appropriate soltanto per il campione che è stato analizzato, che appartiene ad un contesto sociale specifico e non sono generalizzabili a tutte le situazioni.

Sarebbe interessante fare un confronto tra i modi con cui vengono percepiti dagli alunni i sintomi della sindrome di burnout dei professori, emersi da questo studio, e i modi con cui i professori valutano loro stessi. Tale paragone permetterebbe di capire se la percezione dei segnali di stress manifestati dai docenti sono recepiti correttamente dagli alunni.

Riferimenti bibliografici

Fabbris L. (1990), *“Analisi esplorative di dati multidimensionali”*, Cleup Editore.

Bolasco S. (1999), *“Analisi multidimensionale dei dati”*, Carrocci Editore.

Saporta G., Bouroche J.M., (2002) *“Analisi dei dati”*, Clu, Napoli.

Masarotto G. Iacus S.M. (2003), *“Laboratorio di statistica con R”*, McGraw-Hill.

Moshe Tatar, Vered Yahav (1999), *“Secondary school pupils’ perceptions of burnout among teachers”*, British Journal of Educational Psychology

Chiogna M., Dispense per i laboratori del corso di Tecniche statistiche di classificazione, Corso di laurea in Statistica e Gestione delle Imprese, Università degli Studi di Padova disponibile al sito web del corso ai soli iscritti all’insegnamento.

