



Università degli Studi di Padova

*Facoltà di Scienze Statistiche*

Corso di laurea in Statistica e Tecnologie Informatiche

*Analisi della percezione del sé fisico negli adolescenti*

**Relatore:** Prof.ssa Laura Ventura  
**Correlatore:** Prof. Attilio Carraro

**Laureando:** Bisello Andrea

Anno Accademico 2005/2006



# INDICE DEI CONTENUTI

INTRODUZIONE	5
CAPITOLO 1 – analisi dei dati e dei questionari	
• 1.1) Il test Eurofit	7
• 1.2) Il questionario Psdq	12
CAPITOLO 2 – tecniche bivariate e test	
• 2.1) Analisi delle correlazioni nel questionario Eurofit	17
• 2.2) Analisi delle correlazioni nel test Psdq	19
• 2.3) Analisi delle correlazione tra Eurofit e Psdq	21
• 2.4) Verifiche di ipotesi	23
CAPITOLO 3 – Analisi delle componenti principali e analisi cluster	
• 3.1) Breve richiamo dell'ACP	29
• 3.2) Applicazione dell'ACP nel test Eurofit	33
• 3.3) Cos'e' e a cosa serve l'analisi Cluster	36
• 3.4) Applicazione dell'analisi cluster nel test Eurofit	37
• 3.5) Applicazione dell'ACP nel questionario Psdq	39
• 3.6) Applicazione dell'analisi cluster nel questionario Psdq	42
• 3.7) Applicazione dell'ACP nella totalità delle variabili	44
CONCLUSIONI	49
APPENDICE	51
RIFERIMENTI BIBLIOGRAFICI	53



## Introduzione

La tesi desidera analizzare come gli adolescenti percepiscono il sé fisico. Sebbene non esista una definizione univoca e generalmente riconosciuta, in base alle diverse prospettive teoriche sull'argomento, si può affermare che il sé fisico riguarda l'insieme di conoscenze, sentimenti, memorie ed esperienze che l'individuo ha nei confronti del proprio corpo. Secondo gli studi di S. Kretchmar del 1994, il corpo ha raggiunto una tale importanza nella società moderna da dover cercar una propria nuova definizione, arrivando al concetto di "*concezione olistica*". Precedentemente, infatti, secondo l'idea del "*dualismo comportamentale*", il corpo veniva considerato come un qualcosa di "*impegnato nel fare ma non nel pensare*". La nuova idea invece prende il nome dalla parola di origine greca Olos, che significa "Intero", nel senso che il corpo, il movimento ed i simboli non verbali vengono considerati ad un livello paritario rispetto a quello dell'educazione intellettuale. Secondo questo orientamento, la finalità della scuola del terzo millennio dev'essere quella di stimolare competenze cognitive e motorie. Risulta quindi estremamente importante che l'insegnante di educazione fisica senta, percepisca, accetti e valorizzi in modo positivo il corpo, trasmettendo questo valore ai suoi studenti, aumentando l'importanza della corporeità nella scuola, la necessità di un'iniziativa educativa di natura olistica e prevenendo quelle terribili conseguenze che le alterazioni dell'immagine corporea possono provocare a livello psicologico nella persona.

I dati su cui si focalizza questa tesi provengono da una scuola media inferiore della provincia di Vicenza, e riguardano 103 ragazzi, dei quali 50 sono maschi. A tale popolazione di riferimento, omogenea per quanto riguarda lo stile di vita, è stato somministrato un *Questionario per l'auto descrizione del fisico* (PSDQ), avente lo scopo di raccogliere una valutazione soggettiva delle proprie qualità. Le informazioni così ottenute sono state poi affiancate ad una batteria di test ginnici, denominata *Eurofit*. L'origine e lo sviluppo di EUROFIT risale al 1977, in seguito all'incontro tra i Direttori di vari Istituti di Ricerca Sportiva, in cui si rilevò la necessità di valutare l'efficienza fisica e di stabilire dei riferimenti per i ragazzi di età scolare che fossero validi in tutta Europa. Affiancando i risultati del questionario e del test, per ogni studente si hanno pertanto a disposizione complessivamente 22 variabili rappresentanti informazioni sia a livello psicologico che a livello fisico. Lo svolgimento della tesi ruoterà attorno a diverse ipotesi, e per farlo verranno applicate diverse tecniche di analisi. Esempi di ipotesi sono:

- Che percezione hanno gli adolescenti del loro sé fisico?
- In cosa si differenzia la percezione del sé fisico fra maschi e femmine?
- Esiste corrispondenza tra la percezione del sé fisico e l'efficienza fisica?

Il lavoro è suddiviso in vari capitoli. Nel primo capitolo si presentano i questionari e le variabili. I test sono stati somministrati agli studenti nel corso della stessa giornata, per questo non vi sono dati mancanti. Per analizzare le variabili si usano prevalentemente grafici e informazioni di sintesi. Le informazioni ottenute risultano utili nel secondo capitolo.

Nel capitolo due, partendo dal dataset completo, si utilizzano tecniche descrittive bivariate e test per la verifica di ipotesi che permettono di dare già alcune risposte alle tre domande chiave della tesi. Nel terzo capitolo si utilizzano tecniche multivariate come l'analisi in componenti principali e l'analisi cluster, che permettono di far emergere informazioni importanti quali ad esempio dipendenze tra variabili, differenze tra i due sessi o ancora l'individuare la presenza di gruppi di studenti omogenei tra di loro.

## **ANALISI DEI DATI E DEI QUESTIONARI**

In questo capitolo verrà condotta un'analisi preliminare sui dati, di tipo descrittivo. Lo scopo di questa procedura è quello di identificare caratteristiche rilevanti dei dati e capire, osservando grafici e informazioni di sintesi, quali siano le eventuali relazioni tra variabili. Certe suggestioni grafiche possono fornire contributi utili non solo allo studio delle ipotesi di partenza, ma anche alla formulazione di nuove teorie sottostanti al fenomeno in esame. I supporti grafici facilitano la lettura dell'informazione statistica, rendendola maggiormente incisiva e rapida, soprattutto in presenza di molte variabili e di molti dati.

### **1.1) Il test EUROFIT**

Per condurre l'analisi si hanno a disposizione i risultati di un test, chiamato Eurofit, che contiene delle rilevazioni riguardanti le capacità fisiche dei giovani in esame.

Si raccolgono quindi i dati relativi a:

- Sesso : il sesso dello studente.
- Bmi : Indice di massa corporea, valuta sinteticamente lo stato di sotto\normo\sovrappeso.
- Altezza : l'altezza dello studente, in centimetri.
- Peso : in Kg.
- Flessione : Da seduti, bisogna flettersi e cercar di raggiungere i piedi con le punta delle dita. Il numero indica di quanti cm si è superato tale soglia. (0 indica che non si è riusciti a superarla) Più il numero è elevato, più il soggetto riesce a flettersi.
- Busto : il numero di flessioni del busto che il soggetto riesce a fare in 30 secondi. Valuta sia la potenza addominale sia la resistenza.
- Braccia : resistenza muscolare. Si tratta di restare sospesi con la forza delle braccia. (1 punto ogni decimo di secondo), per valutare la resistenza muscolare e la forza.
- Salto: Salto in lungo da fermo. Valuta la forza esplosiva delle gambe.
- Velocità : indica il tempo necessario a fare 5 volte alternativamente un percorso di 5 metri, girandosi, per un totale di 50 metri. Il risultato viene moltiplicato per un fattore 10.
- Equilibrio : un valore basso indica un equilibrio maggiore.
- Battuta : indica il tempo impiegato per svolgere un set di battute.
- Resistenza : indica per quanto tempo l'interessato è riuscito a correre restando nei limiti fisiologici, rilevato da un apposito strumento.

Nella Tabella 1.1, sono riportate le principali informazioni sulle undici variabili continue del test Eurofit. Successivamente verranno forniti i boxplot e i grafici della densità di ogni variabile.

<i>nome</i>	<i>minimo</i>	<i>1 quartile</i>	<i>mediana</i>	<i>Media(sd)</i>	<i>3 quartile</i>	<i>massimo</i>
Bmi	14.02	18.09	20.26	20.68(3.35)	22.56	31.77
Altezza	143.0	156.0	162.0	160.9(7.3)	166.0	176.0
Peso	34.30	46.00	52.10	53.79(10.59)	61.05	82.50
Flessioni	2	12	15	15.41(6.38)	20	30
Busto	4	14.50	18	17.89(4.50)	21	27
Braccia	1	73	138	150.7(124.6)	183	850
Salto	105	145	155	156.6(21.3)	166	225
Velocità	176	204	213	214(16.45)	223.5	270
Equilibrio	0	5	8	9.36(5.79)	14	31
Battuta	84	109.5	120	125(23.83)	137	236
Resistenza	62.5	242	303	310(110.5)	364	602.4

*Tabella 1.1 : Indicatori di sintesi delle variabili del test Eurofit*

Nella Tabella 1.2, invece, sono riportati i risultati del test di Shapiro Wilk per valutare l'approssimazione alla normale delle variabili.

<i>Valutazione oggettiva su...</i>	<i>P-value</i>	<i>Normalità?</i>
Body mass index	0,04	no
Altezza	0,29	sì
Peso	0,03	no
Flessioni del torso	0,02	no
Piegamenti del busto	Circa zero	no
Sospensione con le braccia	Circa zero	no
Salto da fermo	0,1	no
Corsa di velocità	0,06	no
Equilibrio a fenicottero	Circa zero	no
Test di Coordinazione(battuta)	Circa zero	no
Corsa di resistenza	Circa zero	no

*Tabella 1.2 : Verifica della normalità*

Quasi tutte le variabili sono incompatibili con l'ipotesi di normalità e pertanto, per le verifiche di ipotesi condotte nel capitolo 2.4 si ritiene opportuno utilizzare il test di Wilcoxon.

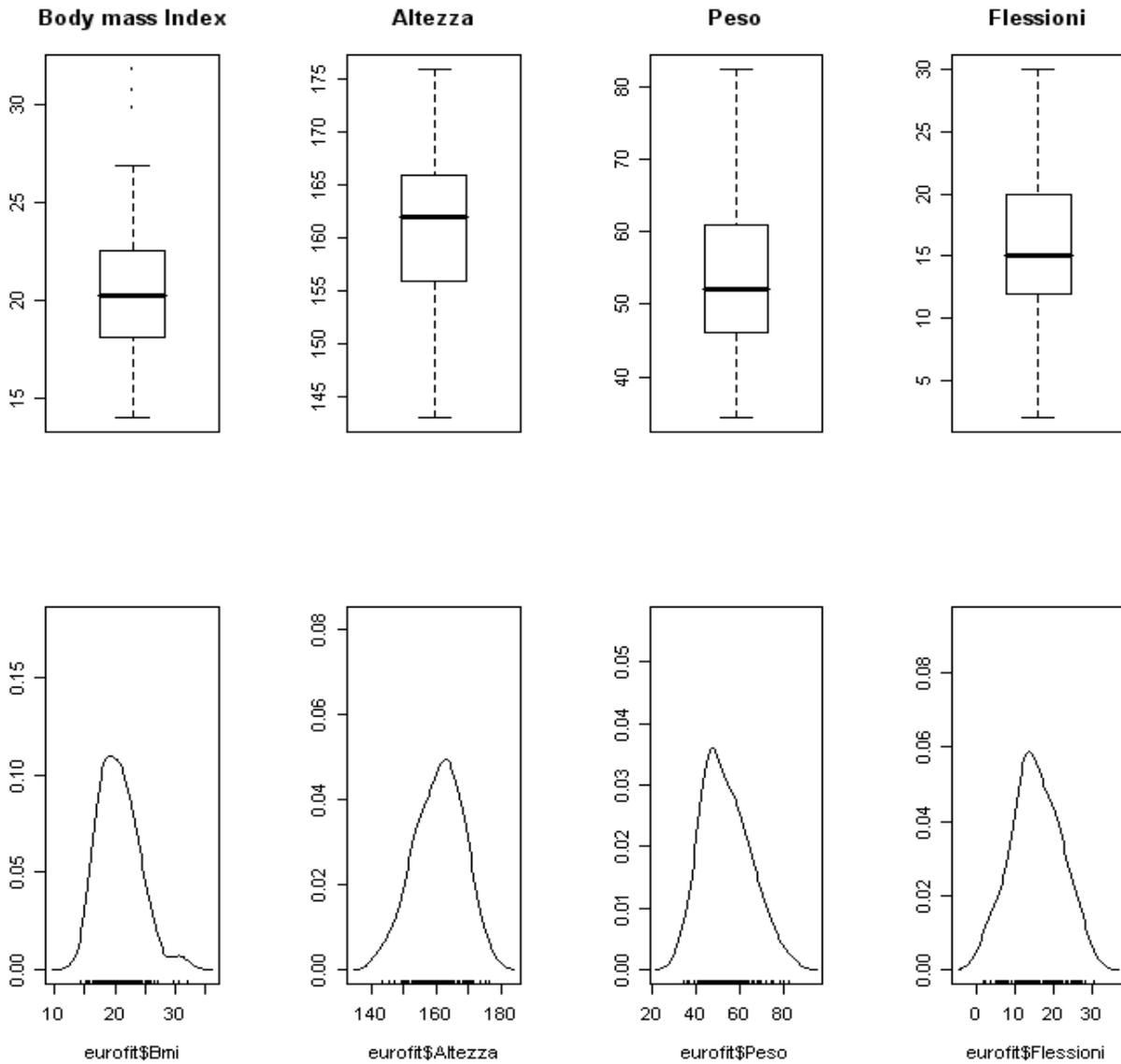


Figura 1.1 : Boxplot e Densità delle quattro variabili sopra indicate

I ragazzi si collocano maggiormente nella fascia di BMI che va dal dal 18 al 23, cioè in quella zona dove, secondo gli standard, si raggiunge il normopeso. Questo rispecchia l'idea che i ragazzi avevano osservando la risposta nella domanda sull'obesità del questionario precedente. Nella coda verso destra sono presenti alcuni valori anomali di Bmi superiore a 30, e quindi in stato di obesità.

L'altezza media degli studenti è di 161 cm per un peso di 53 kg.

Dopo le prime tre variabili di tipo antropometrico, ci sono otto variabili che misurano la prestazione fisica. La prima in esame è il risultato delle flessioni, da non confondere con i piegamenti del busto. Infatti, in questo caso si tratta di dover, da seduti, piegarsi in avanti e oltrepassare la punta dei piedi con le mani, senza flettere le gambe. Il valore rappresenta di quanti centimetri si riesce ad avanzare. La variabilità attorno alla media è notevole: si va dai 2 cm fino ad un massimo di 30 cm, con una media di 15.41 cm. Le quattro variabili che seguono valutano la forza degli addominali, misurata

con flessioni del busto, la forza delle braccia, valutando il tempo di sospensione su un'asta, il salto da fermo, che misura la potenza delle gambe, e la velocità. Per quanto riguarda la velocità, questa variabile valuta un tempo di percorrenza. Un valore elevato potrebbe far pensare ad un individuo veloce, in realtà si tratta di un individuo lento perchè impiega molto tempo per percorrere un determinato tragitto. Di conseguenza, anche questa variabile è stata trasformata, sostituendone i valori con il loro opposto.

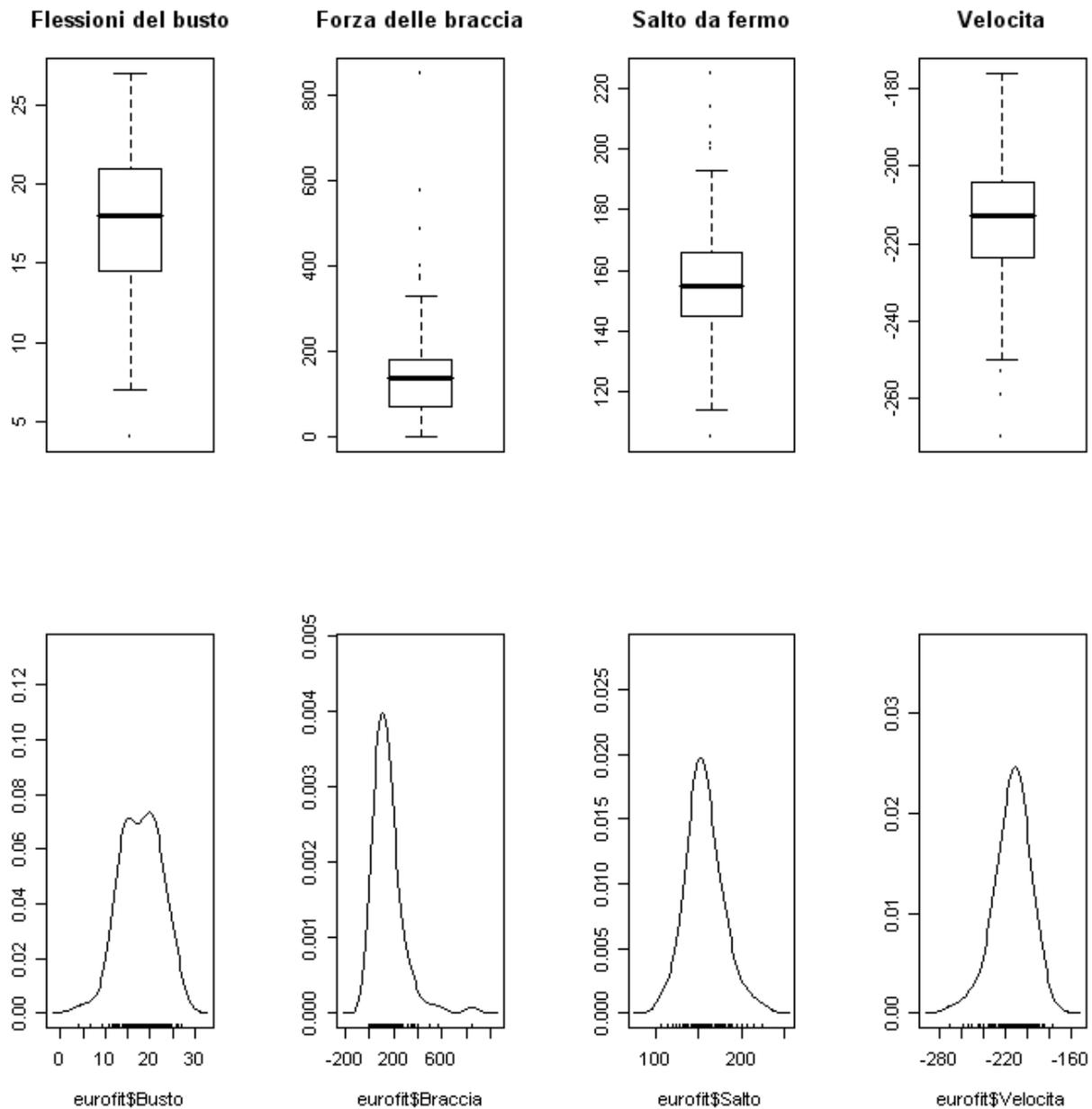


Figura 1.2 : Boxplot e Densita delle seconde quattro variabili sopra indicate

La variabile flessioni del busto valuta il numero di piegamenti che si riesce a fare in un unità standard di tempo. Ovviamente un numero elevato indica un rendimento muscolare maggiore. La media è di circa 18 flessioni. Per quel che riguarda le braccia, vi è una notevole quantità di dati

anomali che spostano di molto la media rispetto alla mediana. La mediana è di 138 secondi. Il salto da fermo misura i centimetri di stacco rispetto al punto di partenza. La simmetria è notevole, infatti media e mediana sono molto vicine, e valgono circa 156 centimetri. Per quanto stabilito negli standard Eurofit, il tempo nella corsa di velocità viene valutato con un numero decimale e poi moltiplicati per un fattore 10 in modo da ottenere un numero intero. In media, per effettuare i 50 metri, servono 21,4 secondi.

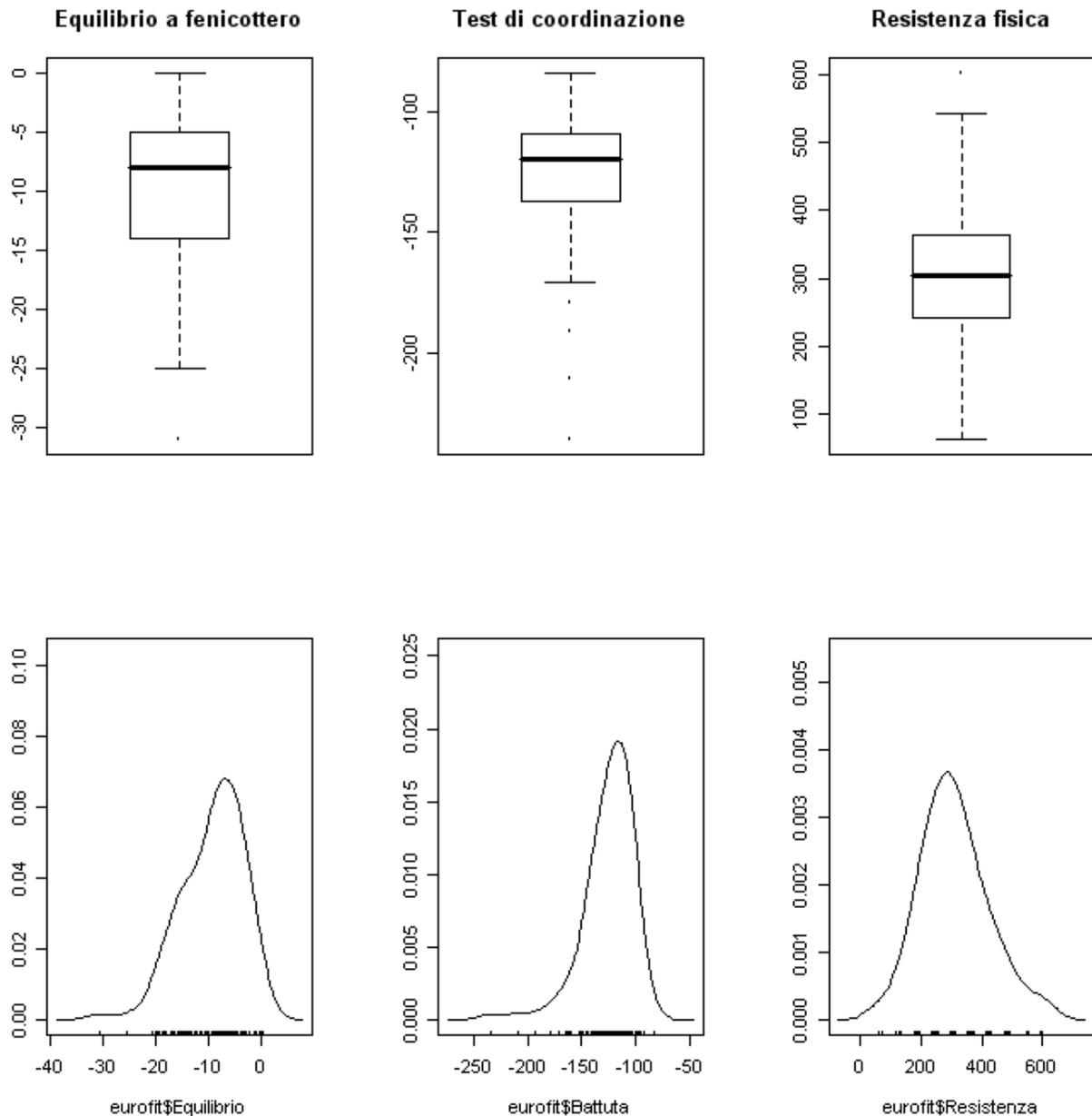


Figura 1.3 : Boxplot e Densità delle ultime tre variabili del test Eurofit

Le ultime tre variabili misurano l'equilibrio, la coordinazione e la resistenza fisica. Per valutare l'equilibrio viene fatto un test chiamato dell'equilibrio a fenicottero. Bisogna restare in equilibrio su un'asticciola sottile. Il risultato valuta il numero di cadute. Anche qui si è deciso di sostituire i dati con il loro opposto, in modo che un valore elevato indichi un buon equilibrio e viceversa. Per la coordinazione si utilizza un test fatto in modo che un tempo di esecuzione basso indichi un'alta

coordinazione. I valori di resistenza fisica, invece, indicano il tempo passato a correre entro certi limiti fisiologici, valutati da uno strumento apposito.

## **1.2) Il questionario PSDQ ( *Physical Self-Description Questionnaire* )**

Il questionario Psdq, costruito per adolescenti dai 12 anni in su, è uno strumento che valuta il concetto fisico del sé nella sua multidimensionalità, attraverso il parere stesso del candidato. Contiene 70 domande che vanno, a rotazione, a rilevare la propria opinione riguardo un aspetto del proprio fisico, dando una votazione da 1 a 6, dove 1 corrisponde al minimo (falso) e 6 corrisponde al massimo (vero).

I valori determinati arrivano da domande come queste:

Salute:

- Quando sono malato mi sento così giù che non riesco ad alzarmi dal letto.
- Mi ammalo così spesso che non posso fare tutte le cose che vorrei.

Coordinazione:

- Mi sento sicuro quando eseguo movimenti coordinati.
- Riesco facilmente a controllare i movimenti del mio corpo.

Attività fisica:

- Più volte alla settimana mi alleno o faccio attività così duramente da respirare a fatica (sbuffare e ansimare).
- Faccio spesso esercizi o attività che mettono duramente alla prova il mio fiato.

Obesità:

- Sono troppo grasso.
- Il mio girovita è troppo largo.

Abilità sportiva:

- Gli altri pensano che io sia bravo a praticare gli sport.
- Sono bravo nella maggior parte degli sport.

Fisico Globale:

- Sono soddisfatto del tipo di persona che sono fisicamente.
- Sono fisicamente soddisfatto di me stesso.

Aspetto:

- Sono attraente per la mia età.
- Ho viso grazioso da guardare.

Forza:

- Sono un individuo fisicamente forte.
- Ho una grande forza nel mio corpo.

Flessibilità:

- Sono molto abile nel flettere, torcere e capovolgere il mio corpo.
- Il mio corpo è flessibile.

Resistenza:

- Posso correre a lungo senza fermarmi.
- Otterrei buoni risultati in una prova di resistenza fisica.

Autostima:

- Nel complesso molte cose che faccio si concludono bene.
- Non ho molto di cui essere orgoglioso. (Reverse Item)

Nel questionario vi è la presenza di alcune *reverse Item*. Per *reverse item* si intende una variabile il cui valore registrato non è il valore inserito dal candidato, ma il valore opposto nella scala di riferimento, secondo la tabella che 1.3. L'uso delle *reverse item* migliora l'attendibilità delle risposte, evitando che il diretto interessato tenda ad auto-influenzarsi.

<i>Valore inserito dal candidato</i>	<i>Valore registrato</i>
1	6
2	5
3	4
4	3
5	2
6	1

Tabella 1.3: Interpretazione delle Reverse Item.

Le variabili presenti sono tutte numeriche e discrete, con valori che vanno da uno fino a sei. Successivamente per ogni variabile viene riportato e descritto il boxplot, che permette di rappresentare graficamente informazioni interessanti sul campione. Nella tabella 1.4 sono stati riportati alcuni indicatori di sintesi. Ad ogni variabile si è assegnato un nome con estensione Psdq per poterla meglio individuare quando verrà formato il dataset completo, che conterrà anche le variabili del test Eurofit. La variabile che tiene conto dello stato di obesità è stata trasformata per poterla meglio interpretare. Ora valori bassi indicano uno stato di obesità basso e viceversa. Nel campione originario, invece, v'era una proporzionalità inversa.

<i>Nome</i>	<i>Minimo</i>	<i>1 quartile</i>	<i>Mediana</i>	<i>Media</i>	<i>3 quartile</i>	<i>massimo</i>
salutepsdq	2.88	4.5	4.88	4.89	5.38	6
coordpsdq	1.67	3.50	4.17	4.16	4.83	6
attivitàpsdq	1	2.33	3.17	3.25	4.17	6
obesitàpsdq	-6	-6	-5.33	-4.72	-3,75	-1
abilitapsdq	1.33	3.50	4.50	4.15	5	6
fisicopsdq	1.17	3.83	4.83	4.58	5.67	6
aspettopsdq	1	3.08	4	3.83	4.83	5.83
forzapsdq	1.5	3	3.83	3.91	4.83	6
fless.psdq	1	2.91	3.67	3.73	4.66	6
resist.psdq	1	2.33	3.17	3.23	4	6
stimapsdq	1.63	4.13	4.88	4.71	5.38	6

Tabella 1.4 : Indicatori di sintesi delle variabili del questionario Psdq.

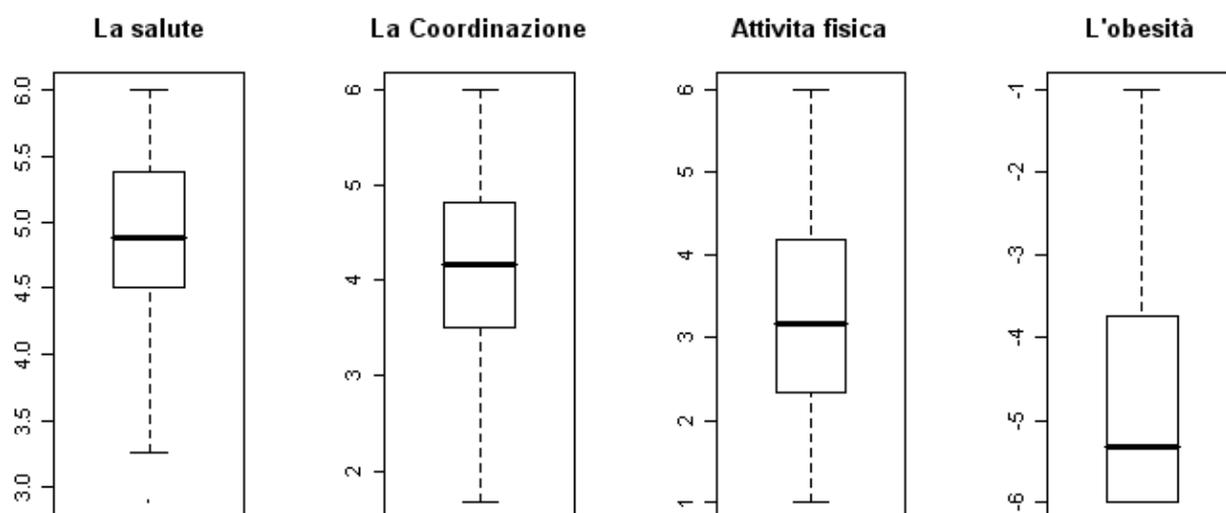
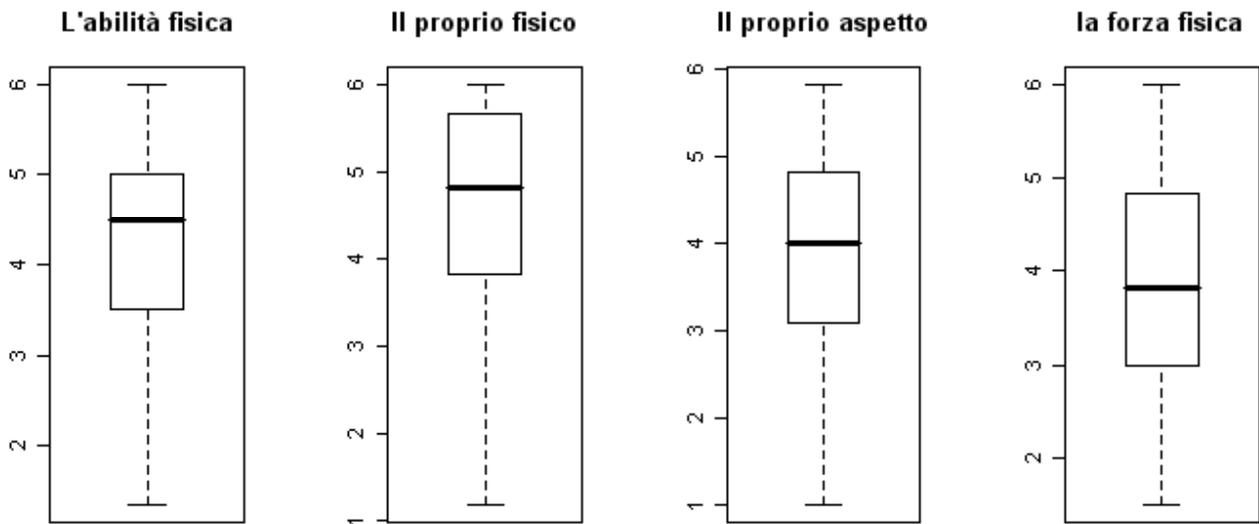


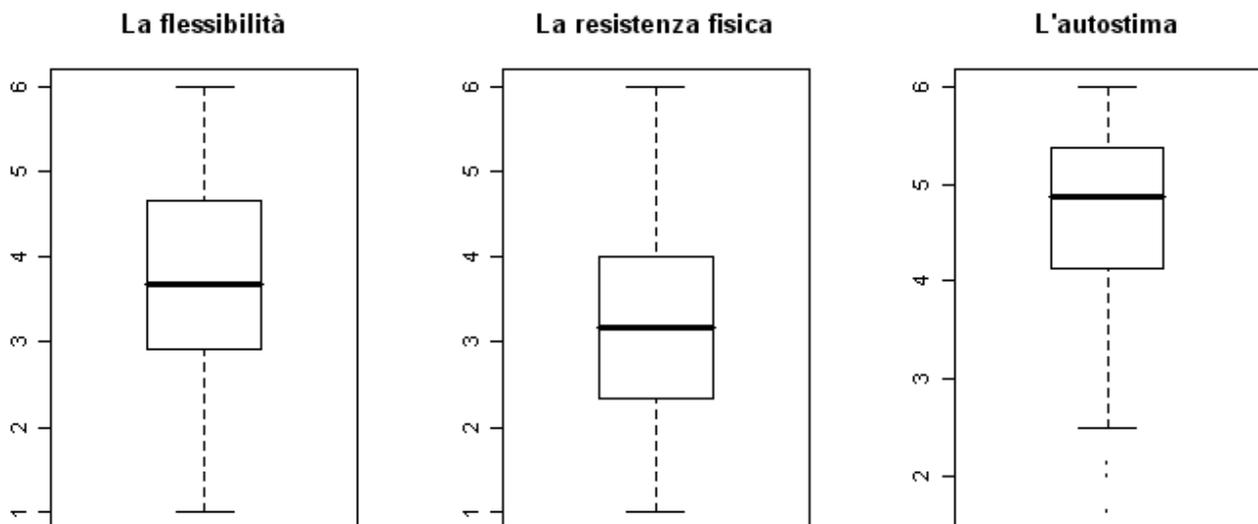
Figura 1.4 : Boxplot delle quattro variabili sopra indicate

Dall'analisi della figura 1.1, si nota come la maggior parte dei 103 ragazzi si vede in buona salute. Infatti, la mediana del campione è 4.88 e la media 4.89. In generale, nel nostro campione, gli studenti si pensano abbastanza coordinati, con un valore medio 4.16. La ripartizione tra i ragazzi che si sentono molto coordinati e coloro che si sentono poco coordinati è buona, data la simmetria attorno al valore centrale. I ragazzi sembrano non essere estremamente interessati all'attività fisica, con un valore medio di 3.25. La scala dei valori sull'obesità è stata rovesciata, per poter ragionare su una proporzionalità diretta. Valori bassi (meno 6) indicano studenti che si pensano non obesi, mentre valori alti (meno 1) indicano studenti che si vedono in forte sovrappeso. Generalmente, in questo istituto, gli studenti sembrano ritenersi non in sovrappeso.



*figura 1.5 : Boxplot delle quattro variabili sopra indicate*

Anche se i ragazzi non sembrano essere estremamente impegnati con lo sport, si ritengono comunque abili nei giochi e negli esercizi ginnici, con una valutazione media di 4.8. La coda sinistra della distribuzione è molto lunga e vi è quindi comunque una discreta percentuale di studenti che si ritengono mediamente abili, o scarsi. La stessa tendenza sembra esserci anche per quanto riguarda l'aspetto fisico, che i ragazzi valutano comunque positivamente. Una controtendenza è presente nella variabile che valuta la forza, dove i valori di media e mediana scendono a circa 3.8.



*Figura 1.6 : Boxplot delle tre variabili sopra indicate*

Le ultime tre variabili del questionario riguardano il sentirsi flessibili, il sentirsi resistenti e l'autostima. Nel caso della flessibilità, vi è simmetria attorno al valore medio 3.73. Lo stesso accade per quanto riguarda la resistenza, dove si nota una variabilità omogenea attorno alla media, anche se con una coda più marcata verso i valori elevati. La maggior parte degli studenti si stima molto, raggiungendo infatti il valore di mediana 4.88.



## Tecniche bivariate e test

### 2.1) Analisi delle correlazioni nel test Eurofit

Allo scopo di individuare la presenza di dipendenze lineari tra le variabili si effettua un'analisi delle correlazioni. Nella Tabella 2.1 sono riportati i valori delle correlazioni calcolate, evidenziando in grassetto le correlazioni significative e segnando con un asterisco le correlazioni non significative.

	Bmi	Alt	Peso	Fles	Busto	Brac	Salto	Veloc	Equi	Batt	Resist
Bmi	1	0,16*	<b>0,89</b>	-0,02*	-0,15*	-0,15*	-0,07*	<b>-0,22</b>	0,04*	-0,05*	<b>-0,30</b>
Alt		1	<b>0,58</b>	<b>-0,19</b>	-0,09*	-0,08*	-0,10*	0,04*	-0,08*	0,04*	0*
Peso			1	-0,11*	-0,17*	-0,16*	-0,02*	-0,17*	-0,01*	-0,02*	<b>-0,26</b>
Fles				1	<b>0,2</b>	-0,06*	-0,14*	-0,17*	0,05*	0,06*	-0,15*
Busto					1	<b>0,4</b>	<b>0,30</b>	<b>0,36</b>	<b>0,25</b>	0,20*	<b>0,39</b>
Brac						1	<b>0,43</b>	<b>0,28</b>	<b>0,25</b>	0,03*	<b>0,48</b>
Salto							1	<b>0,48</b>	<b>0,34</b>	-0,06*	<b>0,52</b>
Veloc								1	<b>0,26</b>	0,15*	<b>0,51</b>
Equi									1	0,13*	<b>0,37</b>
Batt										1	0,06*
Resist											1

Tabella 2.1 : Analisi delle correlazioni del test Eurofit

Nel commentare le correlazioni significative, si nota come il Bmi correli negativamente con la velocità nella corsa e con la resistenza fisica. E' ragionevole pensare che un peso elevato possa gravare sulle gambe determinando un deficit prestazionale, mentre sono proprio risultati elevati negli esercizi che implicano maggiormente l'uso dei muscoli delle gambe a portare ottimi risultati nella resistenza fisica. L'altezza, com'è ovvio, implica una proporzionalità diretta col peso, ma poi non si ripercuote sensibilmente sulle prestazioni fisiche. I ragazzi alti sembrano però essere in difficoltà nel flettere il proprio corpo e ottengono risultati inferiori rispetto ai ragazzi bassi nella valutazione del test di flessibilità. Per quanto riguarda il peso, giungiamo a conclusioni simili all'analisi della correlazione del Bmi, anche se meno intensamente. Infatti, ad un peso elevato può corrispondere anche un'altezza elevata e quindi un mantenimento del Bmi nei valori di normalità. I dati dimostrano come la flessibilità del corpo sia legata principalmente alla potenza degli addominali e non al resto della muscolatura. Vi è un certo legame tra le quattro variabili che determinano la potenza muscolare di gambe, braccia, tronco, che correlano tutte assieme e

significativamente. Buone prestazioni sulle gambe si ripercuotono positivamente sia sul salto in lungo da fermo, sia sulla velocità, permettendo contemporaneamente di resistere a lungo nella corsa di resistenza. Infine, la battuta del disco, come test di coordinazione, non sembra legata a nessun altro aspetto fisico. Di seguito è riportato il grafico della matrice delle correlazioni, dove con un colore più scuro sono segnate quelle significative.

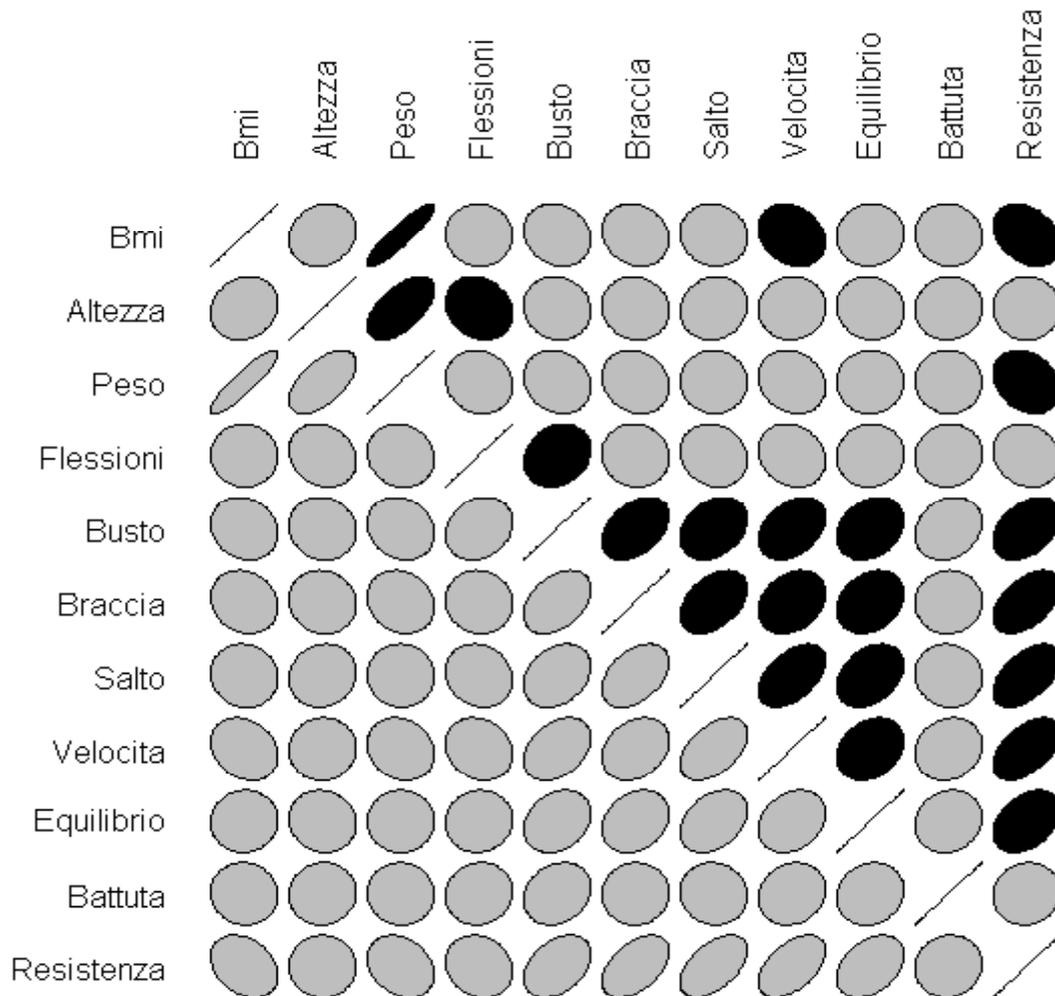


Figura 2.1 : Il grafico delle correlazioni nel test Eurofit

## 2.2) Analisi delle correlazioni nel questionario Psdq

La tabella 2.2 riporta le correlazioni tra le variabili del questionario Psdq. In grassetto sono state evidenziate le correlazioni significative mentre con un asterisco sono state evidenziate le correlazioni non significative.

	salute	coord	attiv	obesit	abilita	fisico	aspet	forza	fless	resist	stima
salute	1	0,07*	-0,03*	-0,02*	0,06*	0,15*	-0,05*	0,14*	-0,05*	0,08*	<b>0,2</b>
coord		1	0,12*	<b>-0,38</b>	<b>0,51</b>	<b>0,57</b>	<b>0,36</b>	<b>0,23</b>	<b>0,72</b>	<b>0,54</b>	<b>0,54</b>
attiv			1	<b>0,31</b>	<b>0,44</b>	0,05*	0,12*	<b>0,32</b>	<b>0,23</b>	<b>0,41</b>	0,05*
obesit				1	-0,07*	<b>-0,56</b>	<b>-0,34</b>	-0,14*	<b>-0,37</b>	-0,18*	<b>-0,40</b>
abilita					1	<b>0,42</b>	<b>0,37</b>	<b>0,46</b>	<b>0,55</b>	<b>0,46</b>	<b>0,49</b>
fisico						1	<b>0,54</b>	<b>0,23</b>	<b>0,46</b>	<b>0,42</b>	<b>0,65</b>
aspet							1	<b>0,27</b>	<b>0,28</b>	<b>0,21</b>	<b>0,57</b>
forza								1	<b>0,27</b>	<b>0,41</b>	<b>0,26</b>
fless									1	<b>0,54</b>	<b>0,41</b>
resist										1	<b>0,33</b>
stima											1

Tabella 2.2 : Analisi delle correlazioni nel questionario Psdq

Tutte queste variabili sono di tipo soggettivo, nel senso che il valore viene attribuito dal candidato stesso. Vi è un forte legame psicologico tra il piacersi e lo stimarsi. Infatti, si nota chiaramente la correlazione significativa di tutte le variabili "positive" con la stima. Una persona che si sente bella, sana, forte, tende ad avere un'elevata autostima di sè. Il vedersi grassi, invece, genera l'effetto opposto, non solo per l'autostima, ma anche nel valutare il proprio aspetto, il proprio corpo, e la capacità di flettersi. Una persona che si vede indifferentemente magra o grassa non cambia però la propria idea rispetto alla forza e all'abilità. Si nota anche un gruppo di correlazioni negative tra l'obesità e il valutarsi efficienti fisicamente, soprattutto per quanto riguarda il proprio corpo e il proprio aspetto fisico. In sintesi, si nota come vi sia una dipendenza forte a livello psicologico nel valutare la propria persona in base a quello che si pensa di essere. Proprio per rispondere a una delle tre domande chiave poste nell'introduzione, bisogna valutare se quanto si pensa di sè è quanto si è veramente. Dovrebbe esserci, quindi, una sorta di dipendenza tra i test Eurofit e il questionario Psdq. Si valuterà se vi è una corrispondenza tale da poter sostenere in che modo gli adolescenti percepiscono il sè fisico.

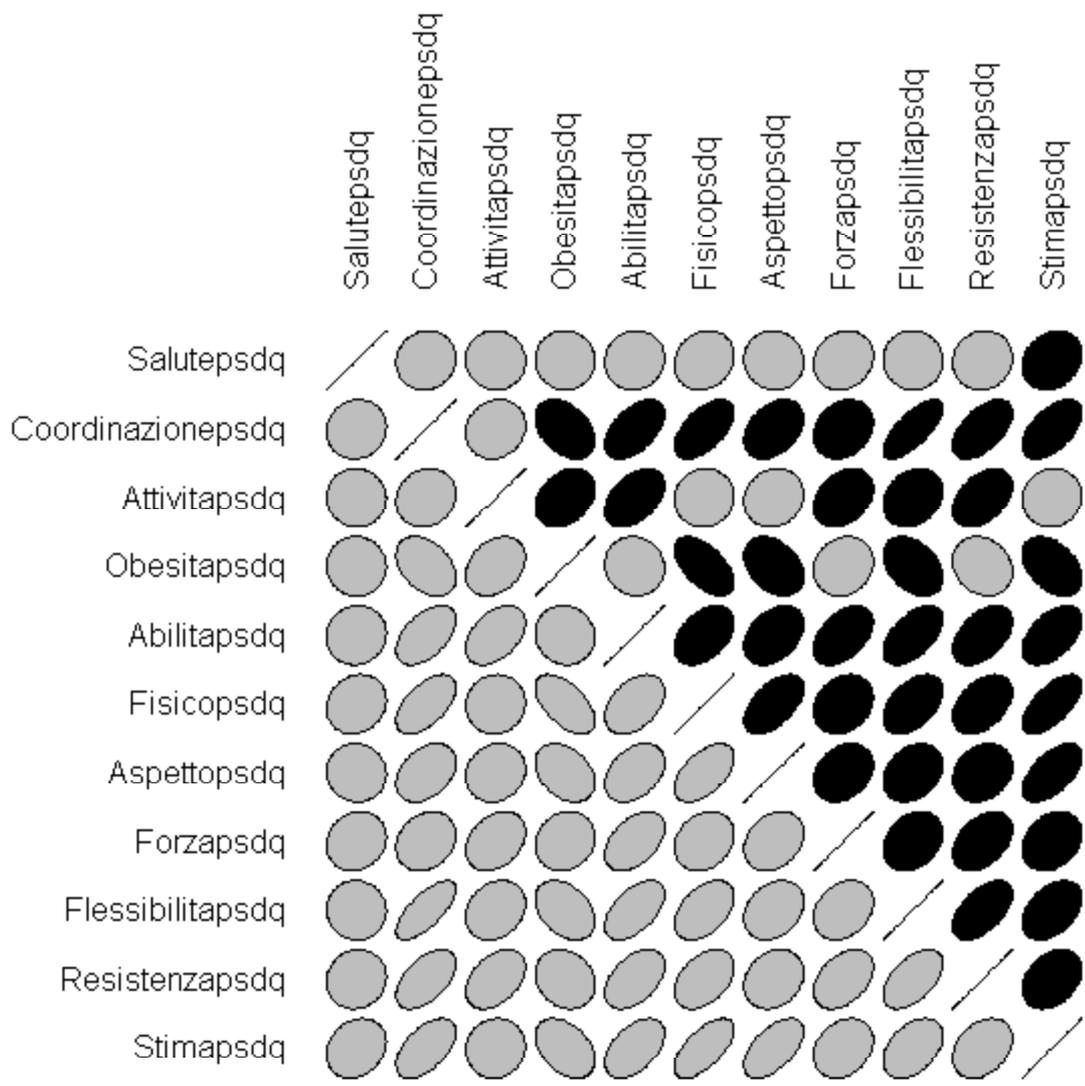


Figura 2.2 : Il grafico delle correlazioni nel questionario Psdq

## 2.3) Analisi delle correlazioni tra Eurofit e Psdq

Nella Tabella 2.3 vengono riportate le correlazioni tra le variabili del test Eurofit e quelle del questionario Psdq. Nella tabella vi sono quindi le correlazioni rimaste, dove in grassetto ho sottolineato quelle significativamente diverse da zero, e con un asterisco quelle che invece possono venir tralasciate.

	Bmi	altezz	peso	flessio	busto	bracci	salto	velocit	Equili	battut	Re ogg
obesita	<b>0,66</b>	0,18*	<b>0,64</b>	0,14*	-0,06*	<b>-0,28</b>	<b>-0,23</b>	<b>-0,24</b>	-0,06*	0,04*	<b>-0,35</b>
forza	<b>0,22</b>	<b>0,30</b>	<b>0,31</b>	<b>-0,10</b>	0,17*	0,18*	<b>0,28</b>	<b>0,26</b>	<b>0,23</b>	0,15*	<b>0,26</b>
salute	-0,02*	-0,11*	-0,06*	-0,01*	<b>0,22</b>	0,19*	-0,05*	0,01*	0,13*	0,12*	0,08*
abilita	-0,18*	0,07*	-0,13*	0,13*	<b>0,34</b>	<b>0,25</b>	<b>0,37</b>	<b>0,44</b>	<b>0,38</b>	<b>0,20</b>	<b>0,50</b>
flessib	<b>-0,37</b>	-0,03*	<b>-0,33</b>	0,05*	<b>0,20</b>	<b>0,32</b>	<b>0,33</b>	<b>0,36</b>	<b>0,22</b>	0,04*	<b>0,44</b>
coordi	<b>-0,34</b>	-0,02*	<b>-0,30</b>	-0,03*	<b>0,21</b>	<b>0,28</b>	<b>0,27</b>	<b>0,37</b>	<b>0,34</b>	0,04*	<b>0,44</b>
fisico	<b>-0,35</b>	0,03*	<b>-0,28</b>	-0,1*	<b>0,21</b>	<b>0,40</b>	<b>0,34</b>	<b>0,33</b>	<b>0,23</b>	0,02*	<b>0,45</b>
Re sog	-0,06*	-0,05*	-0,06*	<b>-0,2</b>	<b>0,30</b>	<b>0,37</b>	<b>0,44</b>	<b>0,25</b>	<b>0,36</b>	0,06*	<b>0,60</b>
attivit	<b>0,21</b>	-0,01*	0,16*	0,07*	0,18*	<b>0,25</b>	0,13*	0,06*	<b>0,26</b>	0,07*	<b>0,36</b>
aspetto	-0,15*	0,04*	-0,11*	-0,02*	0,12*	0,11*	0,08*	<b>0,27</b>	<b>0,27</b>	<b>0,24</b>	<b>0,22</b>
stima	<b>-0,28</b>	0,06*	<b>-0,20</b>	-0,02*	<b>0,27</b>	<b>0,27</b>	0,18*	0,19*	0,18*	0,16*	<b>0,37</b>

Tabella 2.3 : Analisi delle correlazioni tra le variabili del test Eurofit e del questionario Psdq

Nella parte sinistra della tabella troviamo tutte variabili di tipo soggettivo, presenti nel questionario Psdq, e invece in alto tutte quelle di tipo oggettivo presenti nel test Eurofit. Vi è una serie di associazioni preliminari che può essere fatta collegando la sfera psicologica con quella fisica quali, ad esempio:

- la concezione di obesità col valore di Bmi rilevato
- la concezione di flessibilità col risultato del test sulla flessione del busto da seduto
- la concezione di forza fisica con i tre test che valutano la potenza delle braccia, delle gambe nel salto da fermo e nel busto
- la concezione di coordinazione, col test delle battute sul disco
- la concezione di resistenza fisica, con la corsa di resistenza.

Molte altre invece non possono venir confrontate direttamente in quanto col test Eurofit non è possibile quantificare oggettivamente parametri come il grado di salute, la bellezza o la propria stima. E' tuttavia interessante andare a valutare a cosa sono legati questi fattori.

Visto il grado di correlazione estremamente significativo tra obesità e Bmi, si può certamente dire

che un ragazzo che si vede obeso, lo è veramente. Si arriva già ad un primo risultato interessante: valutando le dipendenze tra il BMI e i risultati dei test ginnici si era notato come un peso elevato andava a costituire un problema sicuramente per le gambe, abbassando il risultato di velocità e resistenza, ma non influiva sulle prestazioni della muscolatura del busto. A livello psicologico invece l'obesità crea una sorta di auto limitazione. Infatti, si nota un andamento decrescente anche nei risultati nel salto in lungo e nel test per le braccia, che invece non sono rilevabili valutando il test Eurofit. Un secondo aspetto direttamente confrontabile è la percezione della propria flessibilità con i risultati del test apposito. Per valutare la flessibilità fisica, infatti, secondo gli standard europei l'esercizio adatto consiste nel flettere il busto da seduti e , senza flettere le gambe, oltrepassare la punta dei piedi avanzando il più possibile con le mani. Ovviamente tanto maggiore è l'avanzamento oltre al livello, tanto più una persona dovrebbe risultare flessibile. Nell'analisi condotta i risultati non sono correlati tra loro. Questo può portare a due conclusioni. La prima è che gli studenti non siano in grado di valutare veramente se sono flessibili o meno, e quindi le risposte date nel questionario Psdq non sono legate ai risultati poi ottenuti nell'esercizio; oppure, in questo ambito, il questionario e il test non sono affiancabili, nel senso che quello che gli studenti pensano di dover valutare del sé fisico non corrisponde a quello che poi effettivamente il test della flessione del corpo misura. Inoltre, non vi è differenza significativa nel ritenersi più o meno flessibili in studenti alti o bassi, anche se poi queste differenze si percepiscono chiaramente osservando i risultati. I tre test sulle braccia, sul salto da fermo e sui piegamenti del busto servono a valutare la potenza fisica. La situazione è simile alla precedente, e porta nuovamente ai due casi descritti : o i ragazzi non sanno valutare la propria forza pensandosi forti quando in realtà non lo sono, oppure i test non possono essere affiancati. Effettivamente però nel test Eurofit manca una prova per valutare la potenza esplosiva delle braccia, come ad esempio il lancio del peso. La correlazione è significativa solamente per quanto riguarda il salto da fermo. Nessuna dipendenza certa tra l'autovalutazione della forza e il risultato dei test sugli addominali e sulla resistenza delle braccia. Potrebbe esserci una soggezione nel valutare la propria forza fisica solamente in funzione della potenza delle braccia, e non degli altri muscoli del corpo. Potenza delle braccia che non viene valutata, se non col test della sospensione, che però riguarda più una resistenza muscolare ad uno sforzo prolungato nel tempo. Le domande effettuate nel questionario Psdq candidate alla rilevazione della propria condizione di resistenza fisica tirano in ballo spesso il riuscire a correre a lungo senza stancarsi. Il test per valutarlo nell'Eurofit è esattamente questo; si tratta di dover correre per più tempo possibile senza scendere sotto ad alcuni livelli fisiologici misurati da uno strumento apposito. Il grado di correlazione, infatti, è molto elevato e gli studenti percepiscono correttamente la propria condizione. Questo non succede invece nel test per la valutazione della coordinazione.

## 2.4) Verifiche di ipotesi

A questo punto l'analisi desidera rispondere alla seconda domanda chiave :

- In cosa si differenzia la percezione del sé fisico fra maschi e femmine?

Opportune verifiche di ipotesi possono essere utili per saggiare la possibilità di una differenza significativa nel risultato dei test tra maschi e femmine. Partendo dal presupposto che il questionario Psdq e il test Eurofit sono collegati, una parte delle dimensioni nel test ginnico ha una controparte nel test psicologico. La Tabella 2.4 valuta appunto le dimensioni associabili direttamente

<i>Psdq</i>	<i>Eurofit</i>	<i>Differenza nei due sessi?</i>
Obesita	Bmi	No / No
Flessibilità	Flessioni	No / No
Forza	Braccia / Salto / Busto	Sì / No / Sì / No
Coordinazione	Battuta	No / No
Resistenza fisica	Corsa di resistenza	Sì / Sì

Tabella 2.4 : Verifica delle ipotesi nelle variabili direttamente associabili

Per queste associazioni nel seguito sono riportati i boxplot e la descrizione del risultato della verifica di ipotesi. Vi sono molte altre variabili presenti delle quali si può saggiare la differenza tra maschi e femmine, però non sono direttamente confrontabili tra il test e il questionario e quindi non fanno da supporto alla domanda chiave. Tutte queste altre variabili saranno trattate successivamente e i risultati si trovano nella Tabella 2.5.

Già nell'osservare i boxplot in figura 2.3 , si nota una certa omogeneità nel Bmi tra maschi e femmine, con una varianza maggiore nel sesso forte. Per saggiare l'ipotesi che la differenza sia significativa, si utilizza il test non parametrico di Wilcoxon. Il risultato del test conferma la prima impressione data dall'osservazione dei boxplot : non vi è una differenza significativa tra i due sessi (p-value 0.78). Ragazzi e ragazze si percepiscono correttamente rispetto a quanto determinato dall'indice Bmi, e entrambi i sessi hanno un rapporto peso altezza simile (p-value 0.09). Si può quindi dire che gli studenti percepiscono correttamente il proprio stato di obesità e che non vi è una differenza significativa nel peso dovuto al sesso dell'individuo.

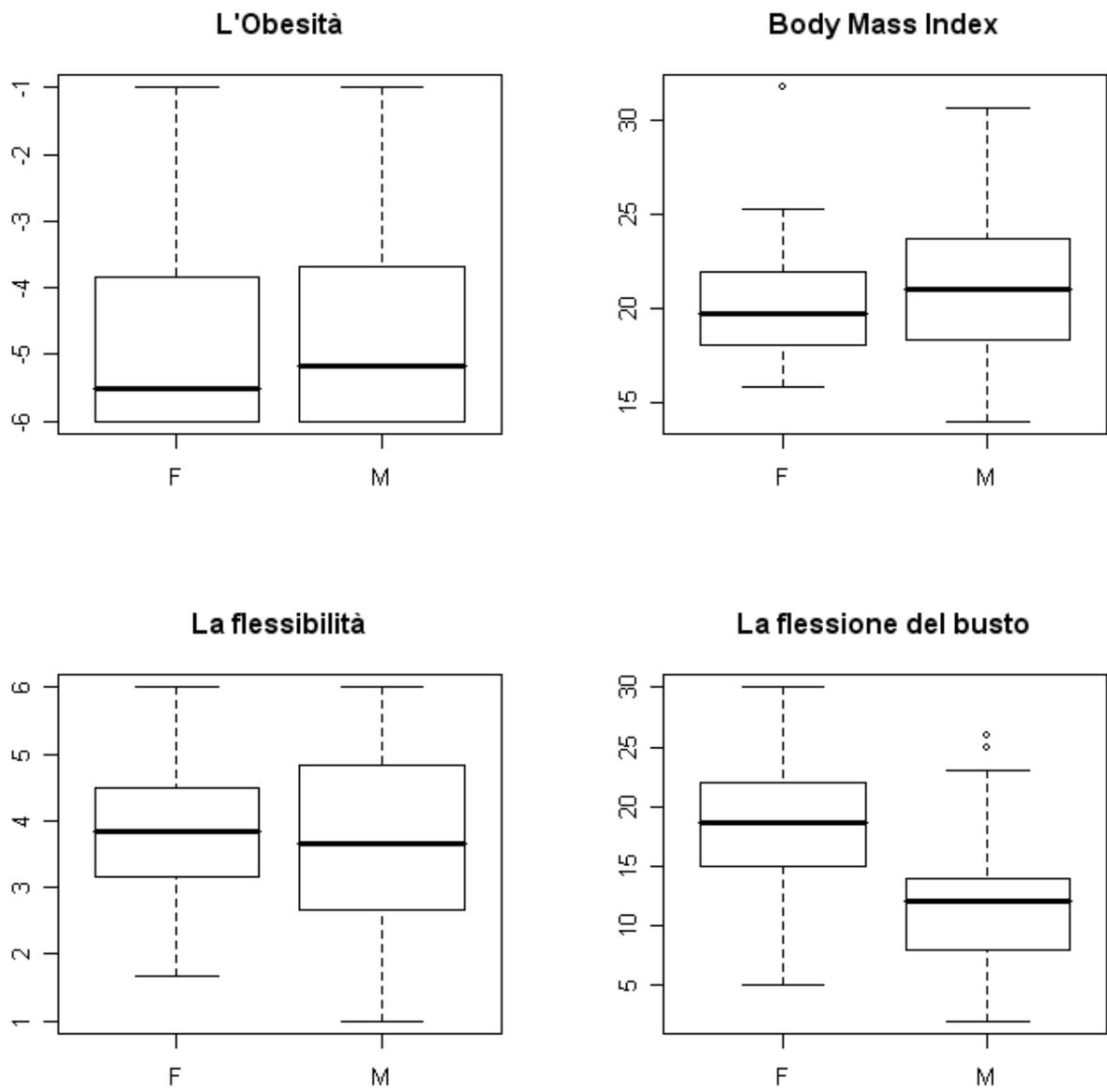


Figura 2.3 : L'obesità e la flessibilità nei maschi e nelle femmine

Per quanto riguarda la flessibilità, è possibile valutare quale dei due sessi si percepisca più flessibile e quale lo sia veramente. Per saggiare la differenza in media tra i due sessi sarebbe possibile usare il test parametrico di student per quanto riguarda la concezione di flessibilità in quanto l'approssimazione con la normale è buona, mentre per i risultati del corrispondente esercizio ginnico si deve usare il test di Wilcoxon. L'impressione data dai boxplot viene confermata dal risultato dei tests : anche se le ragazze sono molto più flessibili dei maschi ( $p\text{-value} \cong 0$ ), psicologicamente non

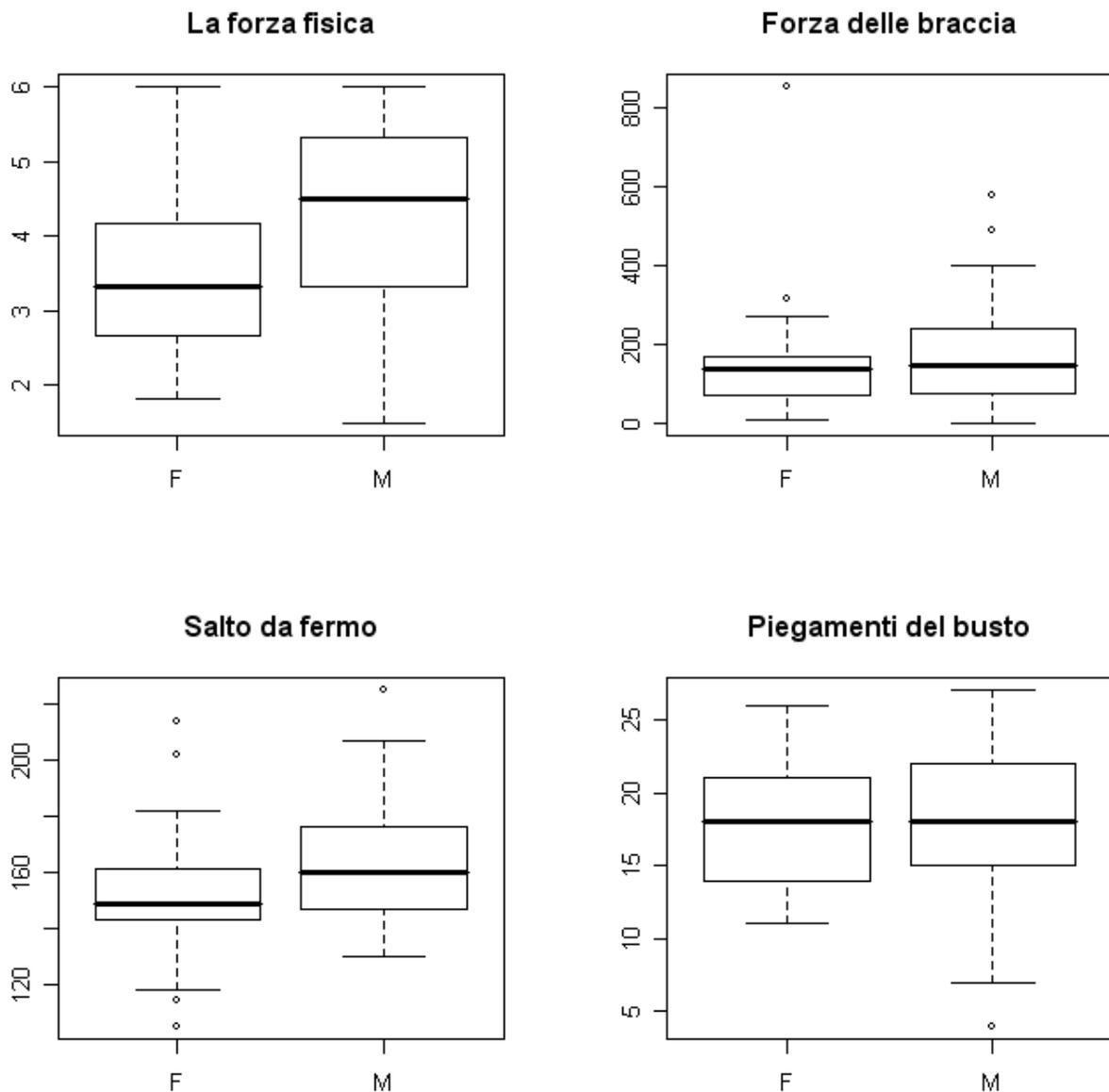


Grafico 2.4 : La forza fisica nei maschi e nelle femmine

si notano differenze ( $p\text{-value} 0.87$ ). I quattro boxplot nel Grafico 2.4 riguardano la forza fisica. I valori riscontrati nel test Psdq vanno a confrontarsi con tre diverse variabili nell'Eurofit. La forza, infatti, viene rilevata dai test sulle braccia, sul salto e sul busto. La differenza per quanto riguarda la valutazione della propria forza fisica nei due sessi è significativa ( $p\text{-value} \cong 0$ ) e i maschi si

reputano più forti. Questo si riscontra solo nel salto da fermo, dove anche c'era correlazione, dimostrando una maggiore potenza muscolare rispetto al gentil sesso. Per quanto riguarda la forza delle braccia e i piegamenti del busto, non vi è motivo di sostenere una differenza tra i due sessi (p-value 0.33 e 0.78). Si notano alcuni valori anomali sia nei maschi sia nelle femmine per quel che riguarda la forza delle braccia nell'esercizio della sospensione e il salto da fermo, con valori leggermente superiori per il sesso forte. I boxplot sono praticamente allineati nei piegamenti del busto, atti a valutare la potenza degli addominali, ma con una varianza maggiore nei maschi. La differenza è significativa nel salto (p-value  $\cong 0$ ), ma non nella forza delle braccia (p-value 0.33). Per quanto riguarda la coordinazione, maschi e femmine oltre a registrare un livello di percezione simile (p-value 0.23), registrano anche risultati allineabili (p-value 0.35). La confrontabilità dei risultati nella resistenza è certa e i maschi oltre a pensarsi più resistenti delle femmine, lo confermano con risultati migliori.

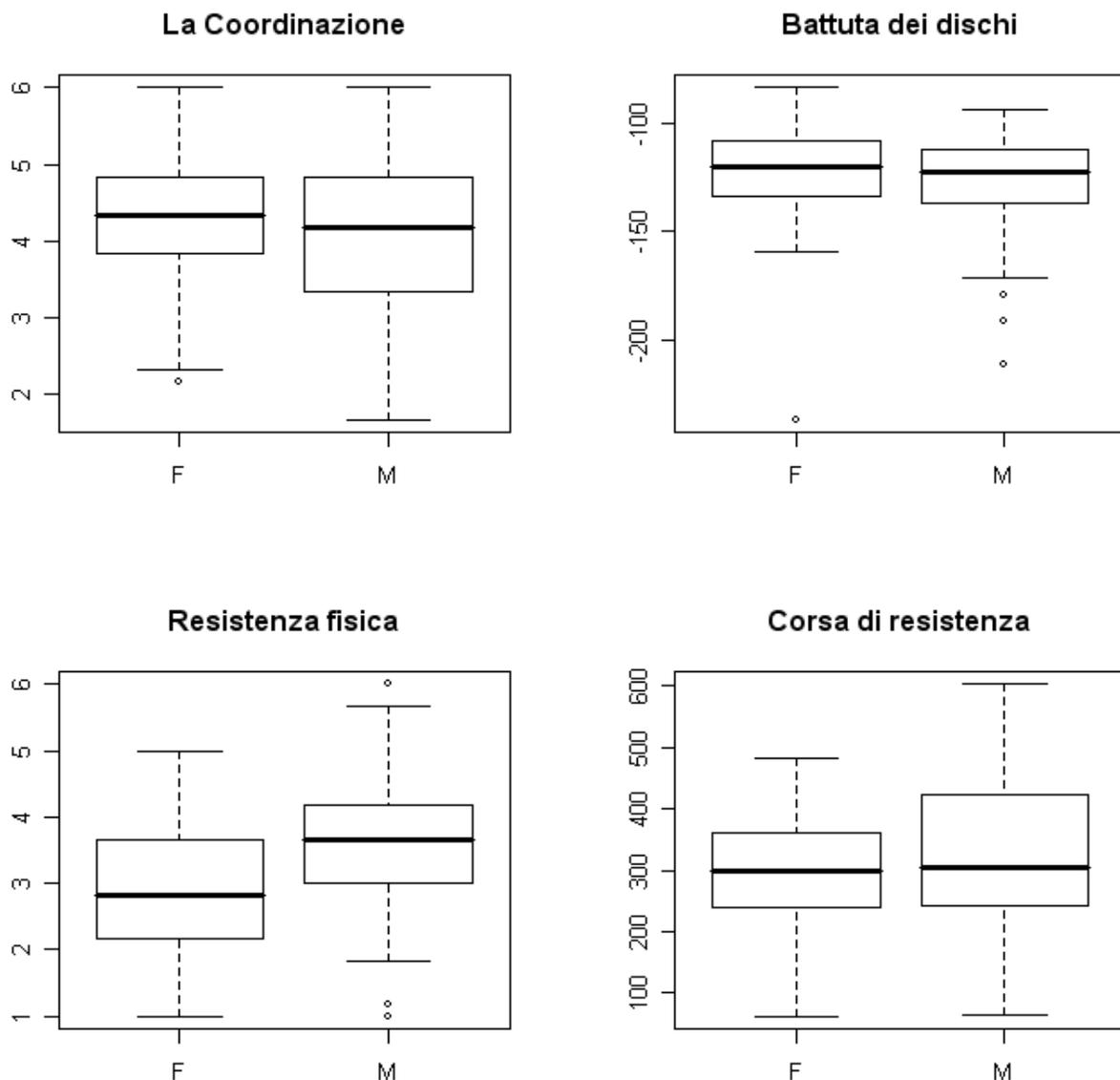


Figura 2.5 : La coordinazione e la resistenza fisica nei maschi e nelle femmine

Anche per tutte le altre variabili, molte delle quali non trovano un confronto diretto dalla sfera psicologica a quella oggettiva, è stata valutata la differenza tra maschi e femmine. I risultati sintetici sono riportati nella Tabella 2.5.

<i>Psdq</i>	<i>Differenza nei due sessi?</i>	<i>Eurofit</i>	<i>Differenza nei due sessi?</i>
Salute	No, p-value 0.25	Bmi	No , p-value 0.09
Coordinazione	No, p-value 0.23	Altezza	No , p-value 0,44
Attività fisica	No, p-value 0.1	Peso	No , p-value 0.13
Obesità	No, p-value 0.78	Flessioni del torso	Sì, Femmine più flessib
Abilità	No, p-value 0.79	Busto	No , p-value 0.78
Fisico	No, p-value 0.53	Forza delle braccia	No , p-value 0.33
Aspetto fisico	No, p-value 0.44	Salto da fermo	I Maschi saltano di più
Forza fisica	Maschi più forti 0.0007	Velocità	No , p-value 0.35
Flessibilità	No, p-value 0.87	Equilibrio	No , p-value 0.68
Resistenza	Maschi più resistenti	Battuta sui dischi	No , p-value 0.35
Stima	No, p-value 0.22	Resistenza	I maschi son più resist

*Tabella 2.5 : Verifica delle ipotesi nelle variabili non direttamente associabili*

Si può notare come nella maggior parte delle variabili provenienti dal questionario Psdq non vi siano differenze tra i due sessi, e le uniche due presenti sono state precedentemente trattate. Nel complesso, le prestazioni fisiche dei due sessi sono simili, anche se il sesso forte sembra prevalere per la resistenza muscolare, mentre il gentil sesso per la flessibilità del proprio corpo.



### Analisi in componenti principali e analisi cluster

L'analisi in componenti principali (ACP) è una tecnica che ha lo scopo di ridurre il numero di variabili in gioco sintetizzandone di nuove, che permettono una più agevole descrizione del fenomeno, cercando di perdere la minor quantità possibile di informazione. *(Per una trattazione dettagliata dell'ACP si rimanda al libro "Analisi esplorativa di dati multidimensionali" del Dott. Fabbris del 1990).*

Nella nostra analisi abbiamo a disposizione 22 variabili e studiarne tutte le possibili combinazioni risulta essere molto complicato. Queste variabili trattano argomenti diversi, ma sono legate tra di loro almeno concettualmente. È lecito pensare sia possibile riuscire a ridurre il numero di variabili in gioco, in quanto alcune di esse potrebbero portare le stesse informazioni e contenere ridondanze, che complicano i calcoli e confondono l'interpretazione dei risultati. L'analisi in componenti principali permette proprio di ovviare a questo problema, creando delle nuove variabili, dette latenti, che riescono in qualche modo a portar lo stesso contributo informativo delle altre. Ovviamente si deve cercare un compromesso tra la riduzione del numero di variabili e la perdita informativa subita.

Segue ora una sintesi del procedimento matematico che sta dietro all'analisi ACP, ma per una spiegazione dettagliata si rimanda comunque al libro di testo sopra citato.

#### 3.1) Breve richiamo dell'ACP

Partiamo dai concetti base di Autovalore e Autovettore. Sia  $X$  una matrice quadrata di dati di ordine  $p \times p$ , e sia  $\underline{w}$  un vettore di  $p$  elementi.

Se viene soddisfatta la condizione

$$\underline{X}\underline{w} = a \underline{w} \quad (1)$$

per qualche scalare  $a \in \mathbb{R}$ , allora  $\underline{w}$  è un autovettore di  $X$  e  $a$  è il suo corrispondente autovalore.

Praticamente l'autovettore di una matrice quadrata è quel particolare vettore che moltiplicato per la matrice di riferimento ha come risultato il vettore stesso al netto di una sua opportuna riscalatura.

Questa riscalatura è ottenuta attraverso l'autovalore  $a$ .

Per calcolare autovalori e autovettori, supponiamo di partire dall'equazione (1) e sottraendo a  $\underline{w}$  da entrambi i termini si ottiene

$$\underline{X}\underline{w} - a\underline{w} = \mathbf{0}, \quad (2)$$

da cui raccogliendo  $\underline{w}$  si ha

$$(\mathbf{X} - \mathbf{a} \mathbf{I})\underline{w} = \mathbf{0}. \quad (3)$$

Eliminando il caso banale in cui  $\underline{w} = 0$ , allora dev'essere  $|\mathbf{B}| = |(\mathbf{X} - \mathbf{a} \mathbf{I})| = 0$ .

L'equazione  $|(\mathbf{X} - \mathbf{a} \mathbf{I})| = 0$  viene detta equazione caratteristica di  $\mathbf{X}$ . La soluzione di questa equazione comporta la ricerca di  $p$  valori di  $a$ , detti radici caratteristiche, o autovalori.

Ordinando gli autovalori non nulli dal più grande al più piccolo e usandoli nella formula (3), si ottengono altrettanti valori  $w$  (autovettori) di coefficienti della prima componente principale.

Per il calcolo delle componenti successive alla prima si esegue lo stesso procedimento dopo aver calcolato una matrice di varianza – covarianza residua, depurata dalla variabilità e covariabilità estratte dalla precedente componente.

Il calcolo di questi valori dal punto di vista statistico è importante nell'analisi delle componenti principali, nella quale si determinano autovalori e autovettori della matrice di covarianza di  $\mathbf{X}$ .

Geometricamente la matrice  $\mathbf{X}$ , che ha  $p$  variabili, può venir rappresentata in uno spazio a  $p$  variabili. Lo scopo è quello di riuscire a proiettare gli  $n$  punti in un sottospazio a  $q$  ( $q < p$ ) variabili fatto in modo che la nuvola di punti proiettata venga "deformata" il meno possibile. Si assume che queste nuove variabili sintetizzate siano una combinazione lineare delle variabili di partenza, una combinazione che deve portare alla varianza massima, e quindi al massimo contributo informativo. La prima componente principale è caratterizzata da una propria varianza, maggiore rispetto alle successive, in ordine. La seconda componente ad esempio, depurata dalla variabilità spiegata dalla prima, è quella che spiega più varianza possibile e così via. La varianza totale viene in pratica ridistribuita nelle componenti principali, e la percentuale di varianza spiegata da una componente è calcolabile come il rapporto tra l'autovalore relativo e la somma di tutti gli autovalori. Ovviamente si cerca un compromesso tra varianza totale spiegata e numero di variabili artificiali ottenuto.

Nel momento in cui si va a generare una componente principale "y", si cerca la combinazione lineare delle variabili che permette di ottenere la maggior varianza possibile.

Si dimostra che questo avviene quando come coefficienti delle variabili di cui si cerca la combinazione lineare si usa l'autovettore "a" associato al più grande autovalore determinato.

Tale procedimento si può ripetere per  $q$  volte, ma risulterà sempre  $\text{var}(y_1) > \text{var}(y_2) \dots > \text{var}(y_q)$ .

Nel voler fare un esempio pratico, si suppone di voler riconoscere un oggetto tridimensionale ( 3 dimensioni, 3 variabili ) dalla sua ombra ( 2 dimensioni, 2 variabili ). Si può ruotare l'oggetto in modo da ottenere ombre diverse di diversa estensione. Quasi sempre l'ombra di massima estensione è quella che può far meglio comprendere la natura dell'oggetto. Ad esempio, supponiamo di mettere un pettine davanti ad una fonte di luce perpendicolare ad essa. Dall'ombra così ottenuta, equivalente alla massima estensione, si riuscirà a capire chiaramente che si tratta di un pettine. Ma se giriamo il pettine per lungo, l'ombra risulterà un piccolo segmento nero, e non sarà possibile capire di cosa si tratta. Ci sono anche delle eccezioni, che implicano l'importanza di avere un

metodo per la rotazione della proiezione. Supponiamo di vedere un messicano con un grande sombrero e una pistola in mano. Visto dall'alto, la proiezione sarà di dimensione massima, ma il cappello quasi sicuramente andrà a coprire particolari importanti (la pistola) che non potremo notare. Guardando il messicano di fronte, abbiamo un'immagine di dimensione minore, ma possiamo notare particolari importanti che prima non sarebbero stati visibili. Ci sono metodi diversi per effettuare la rotazione e comunque ruotando in qualsiasi maniera la proiezione dei punti la varianza spiegata non cambia.

In linea di principio vi sono metodi di rotazione ortogonali e metodi di rotazione obliqui. La differenza fondamentale tra i due è che i primi soddisfano la condizione di ortogonalità tra gli assi fattoriali, mentre i secondi no.

Le principali tecniche di rotazione sono Varimax, Quartimax, Oblimin e Promax.

- Varimax : aumenta la semplicità dei fattori massimizzando la varianza delle saturazioni, quindi porta a 1 le saturazioni già elevate e a 0 quelle basse.
- Quartimax : tende a formare un unico fattore generale.
- Oblimin : crea una rotazione obliqua, fa in modo che le variabili abbiano saturazioni vicine a 0 in tutti i fattori tranne uno.
- Promax :un'altra rotazione obliqua, parte da una rotazione Varimax, poi la modifica per renderla più semplice aumentando i pesi fattoriali già elevati e riducendo quelli più piccoli, permettendo la correlazione tra i fattori.

Utilizziamo la tecnica dell'analisi in componenti principali per scoprire nuove caratteristiche nel dataset. La tecnica verrà utilizzata sia nelle undici variabili di tipo soggettivo ricavate dal questionario Psdq, sia dalle undici oggettive ricavate col test Eurofit, e poi la stessa tecnica verrà utilizzata nella fusione di tutte. L'analisi avviene in tre parti:

1. Scelta del numero di componenti principali.
2. Estrazione dei fattori.
3. Interpretazione degli assi.

Per quanto riguarda il numero di componenti principali, vi sono 4 linee guida:

1. Il numero viene scelto a priori. Questo metodo si usa quando si ha già una certa conoscenza del fenomeno.
2. Il numero di fattori viene scelto in base a un valore minimo di varianza spiegata. Una percentuale del 75% viene considerata un buon traguardo.
3. Si prendono solo gli autovalori maggiori di 1 (Guttman 1954, Kaiser 1960)

4. Metodo del punto di gomito : permette di individuare gli autovalori importanti meglio di quanto sia possibile con la lettura dei valori stessi (Fabbris 1990). Si rappresentano gli autovalori sul piano cartesiano e si collegano con una spezzata. Se si evidenzia chiaramente una forte inclinazione nei primi fattori e un appiattimento nei successivi allora si possono ignorare gli autovalori successivi al cosiddetto punto di gomito, cioè dove la spezzata cambia rapidamente inclinazione.

Una volta scelto il numero di componenti da estrarre, queste vengono costruite come precedentemente illustrato. Successivamente si deve trovare un'interpretazione ragionevole al significato degli assi, che rappresentano le componenti principali.

Dare una valida interpretazione può risultare molto difficile, ma è la conclusione dell'analisi ACP. Le variabili nel cerchio delle correlazioni vengono identificate come delle frecce. Le frecce più lunghe e che più si avvicinano alla circonferenza sono quelle che più correlano con i fattori, e che vanno maggiormente studiate per riuscire a identificare il significato del fattore. Le frecce che non si allontanano molto dal centro, hanno anche un'importanza minore. L'ideale sarebbe avere variabili molto prossime alla circonferenza vicine al centro.

### 3.2) Applicazione dell'ACP nel test Eurofit

Si inizia valutando il numero di fattori da estrarre, con la tecnica del punto di gomito prima, e della varianza spiegata dopo.

Il grafico che segue mostra l'andamento degli autovalori calcolati per le variabili nel test Eurofit.

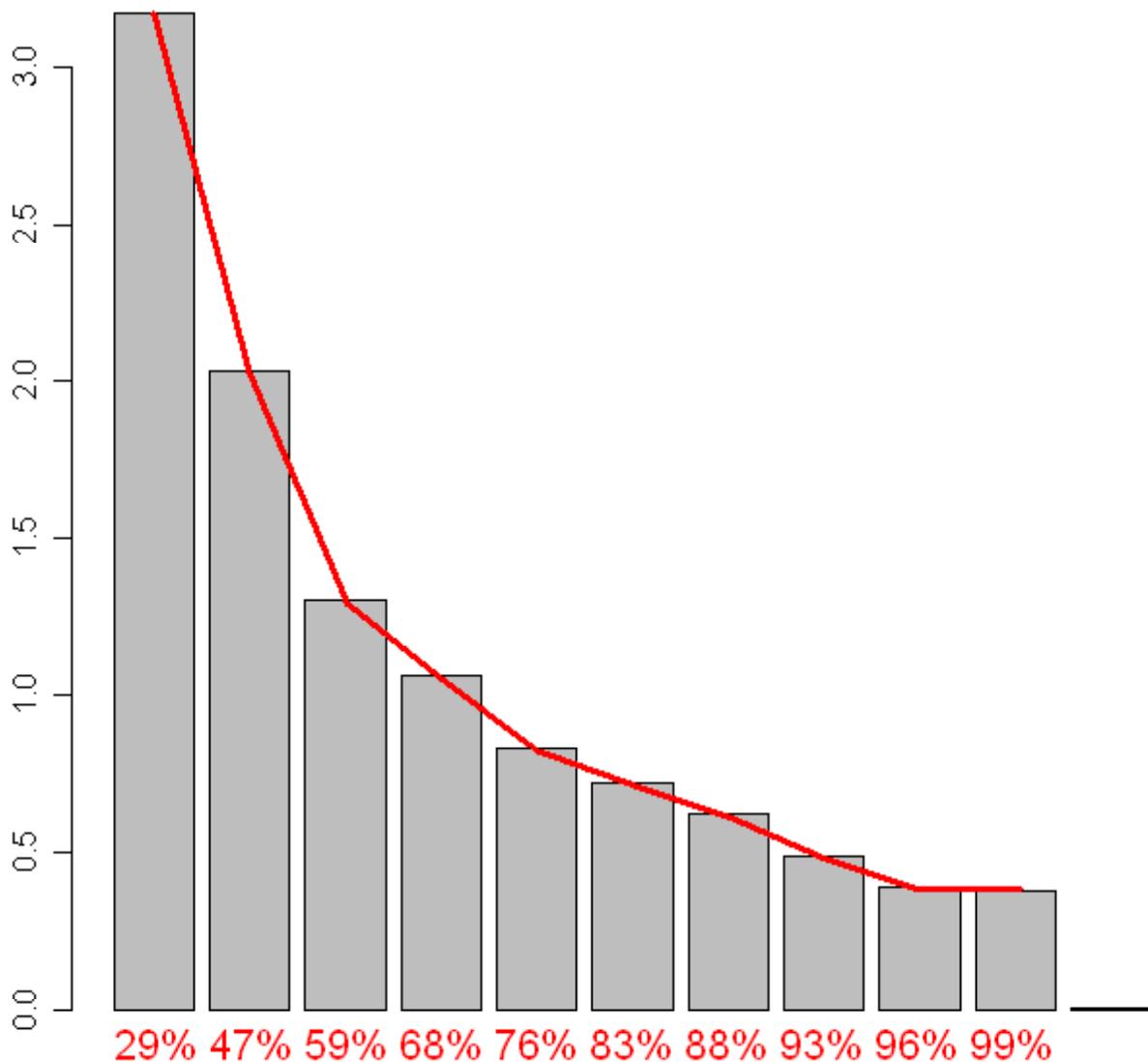
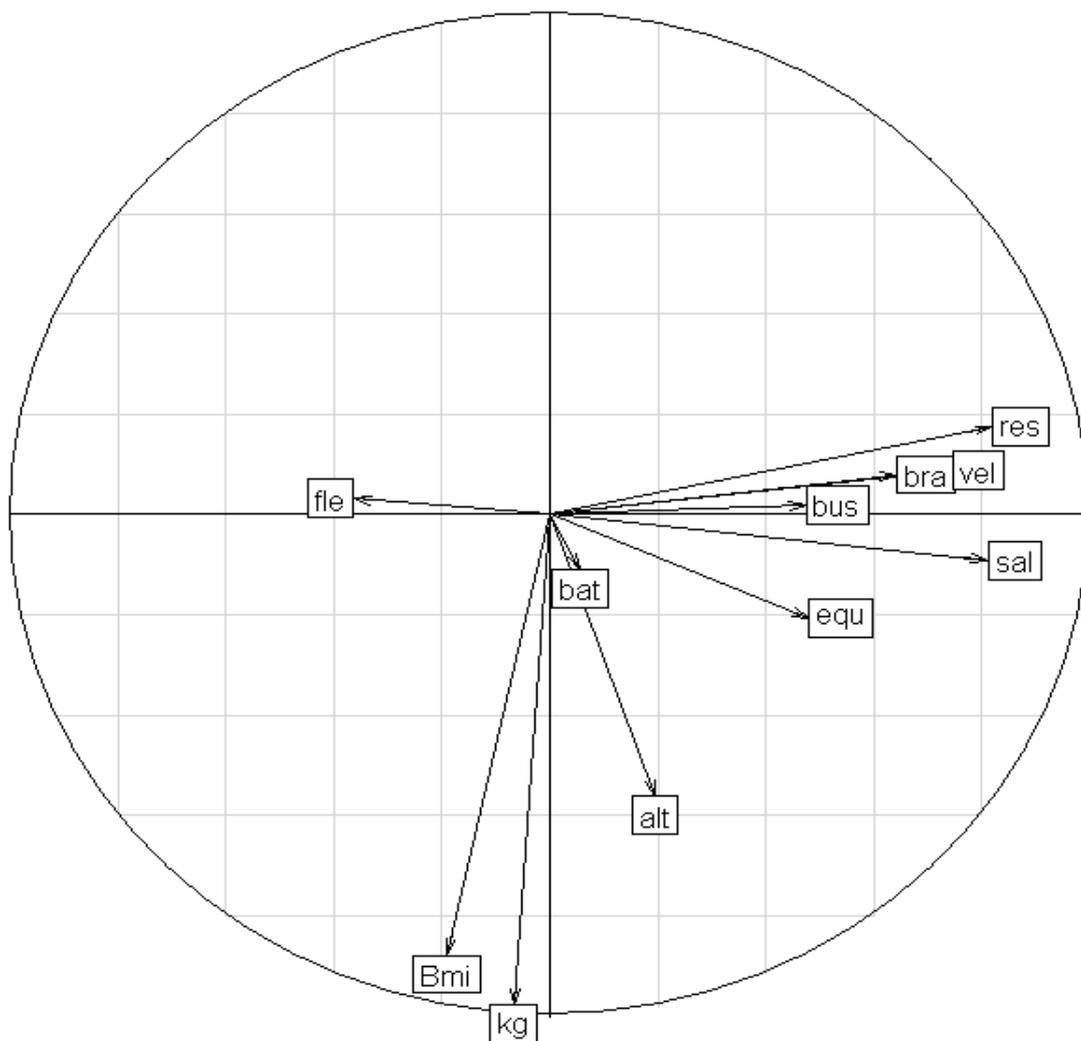


Figura 3.1 : Autovalori nell'analisi del test Eurofit

Si ritiene opportuno l'utilizzo di due fattori, che riescono a spiegare rispettivamente il 47% della varianza totale. Per migliorare la leggibilità del cerchio sono state rinominate le variabili secondo quanto riportato nella Tabella 3.11. Successivamente segue il cerchio delle correlazioni (Figura 3.2).

<i>Variabile originale</i>	<i>Variabile rinominata</i>
Bmi	Bmi
Peso	Kg
Altezza	alt
Flessioni	fle
Battuta (Coordinazione)	coo
Equilibrio	Equ
Salto	sal
Velocita	vel
Braccia	bra
Resistenza	res
Busto	bus

*Tabella 3.1 : Rinominazione delle variabili*



*Figura 3.2 : Cerchio delle correlazioni*

Osservando il cerchio delle correlazioni col primo e il secondo fattore si arriva a conclusioni interessanti, che confermano in parte già quanto spiegato precedentemente. Intanto, è possibile attribuire all'asse verticale il significato di **misure antropometriche** e all'asse orizzontale il significato di **potenzialità fisica**. Si nota come non vi sia dipendenza tra l'altezza o il peso di un individuo e le relative potenzialità fisiche. Questo si notava anche con la matrice delle correlazioni, dove ad un Bmi elevato si trovavano correlazioni significative solo con la resistenza e la velocità. La battuta e le flessioni, vista la vicinanza al centro del cerchio, non forniscono informazioni importanti. Anche il cerchio tra il secondo e il terzo fattore porta allo stesso risultato, con le variabili antropometriche nell'asse orizzontale e quelle fisiche nell'asse verticale.

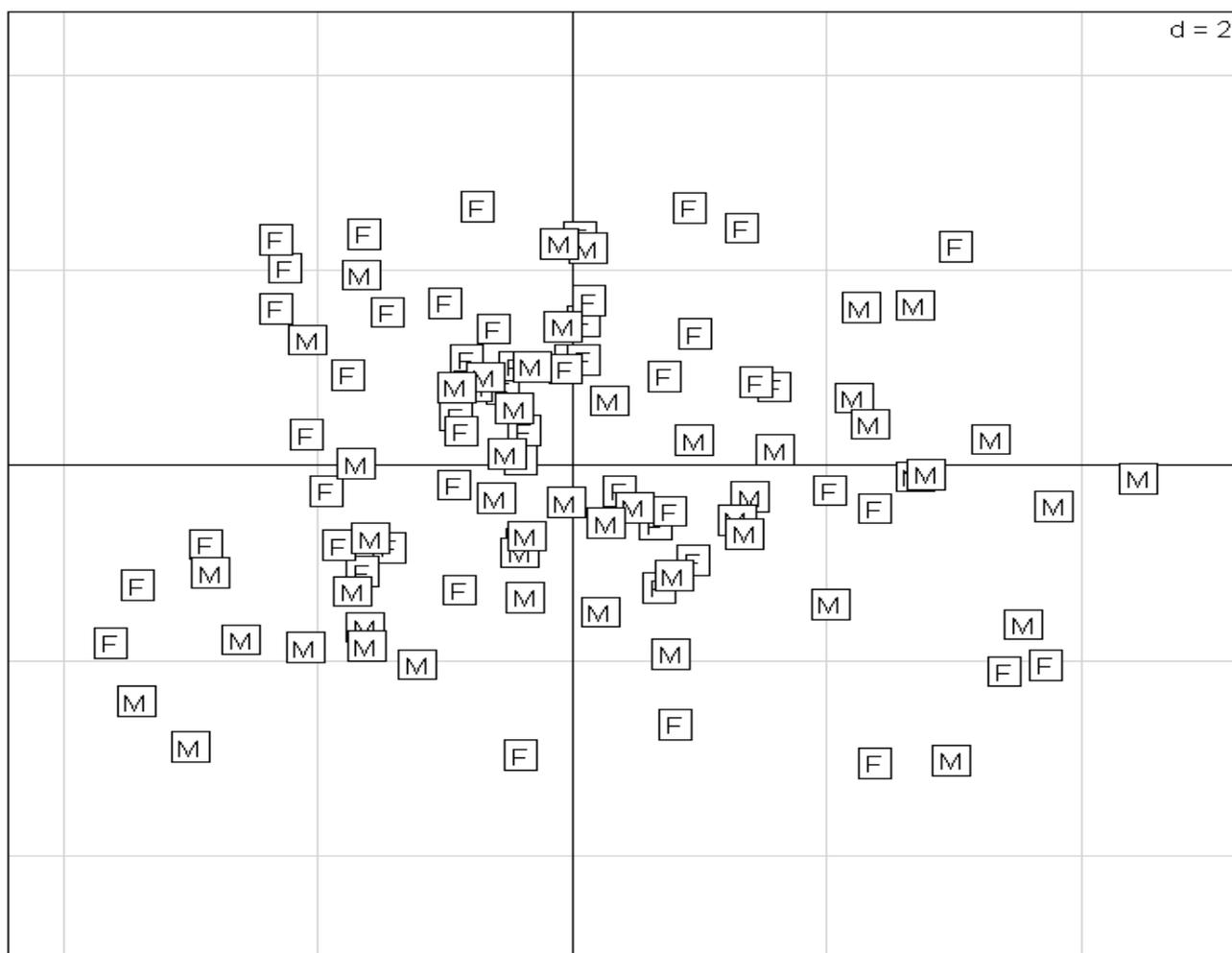


Figura 3.3 : Punteggi fattoriali

Osserviamo la rappresentazione grafica dei punteggi fattoriali. Si nota che la popolazione di riferimento non tende a formare particolari gruppi, e tutti i candidati spaziano attorno al centro, sia maschi che femmine, senza formare gruppi compatti in relazione al sesso. Più ci spostiamo verso destra, più otteniamo prestazioni fisiche elevate, mentre più ci spostiamo verso il basso, più troviamo i ragazzi pesanti. E viceversa. Si può usare il dendrogramma come tecnica di raggruppamento, alla ricerca di gruppi omogenei tra loro. Nel paragrafo seguente verrà illustrata la tecnica utilizzata.

### 3.3) Cos'è e a cosa serve l'analisi Cluster

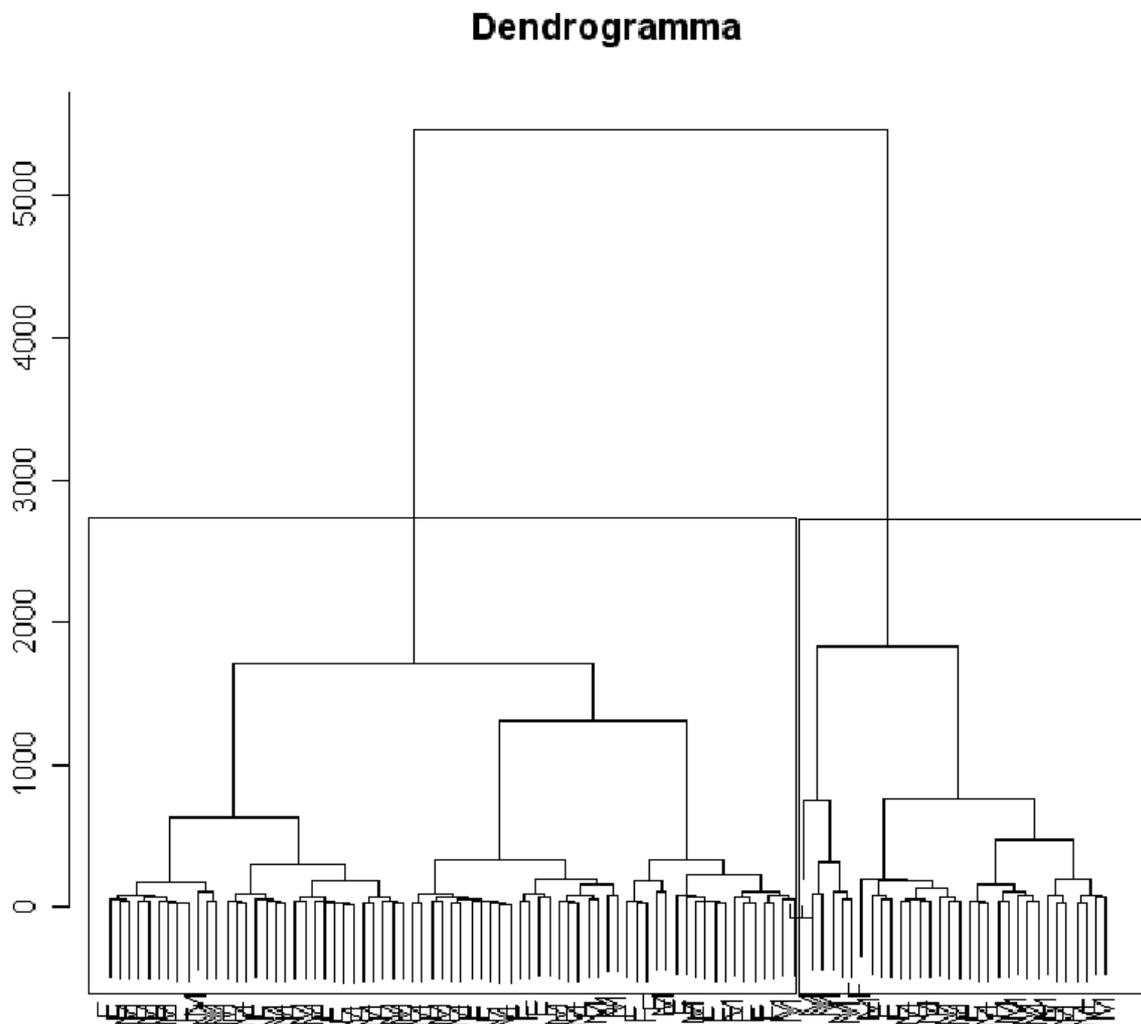
Lo scopo dell'analisi Cluster è quello di riuscire a raggruppare  $n$  unità disponibili in  $k$  gruppi. Non esiste una variabile risposta e tutte quelle disponibili hanno la stessa importanza. Un utilizzo tipico è quello della segmentazione della clientela da parte di un'azienda che se riesce a identificare gruppi di clienti omogenei tra loro, può pensare ad azioni di marketing mirate. Uno degli aspetti più complicati del clustering è trovare il giusto metodo per valutare la "vicinanza" o la "lontananza" delle unità. Vi sono molte tecniche che trovano più o meno utilizzo a seconda del tipo di variabili utilizzate. Vi sono metodi gerarchici e non gerarchici. Il metodo non gerarchico più noto è quello del K-Medie, che si utilizza in caso di variabili continue ma ha il grosso inconveniente di dover specificare a priori quale sia il numero di gruppi che si vogliono formare. Nelle tecniche gerarchiche troviamo un'ulteriore suddivisione, in tecniche agglomerative e non agglomerative. Un metodo agglomerativo procede partendo dall'idea che ogni unità è un gruppo a parte e forma gruppi man mano sempre più grandi unendo a loro volta i gruppi in qualche modo più "vicini". Il metodo maggiormente usato per misurare la distanza fra i casi è la **distanza Euclidea al quadrato** definita come la sommatoria delle distanze al quadrato fra tutte le variabili di due differenti gruppi:

$$\text{Distanza}(x, y) = \sum_i (x_i - y_i)^2$$

Nell'analisi condotta viene utilizzato il metodo gerarchico agglomerativo di Ward per la formazione dei gruppi. Il metodo di Ward combina i cluster in modo tale che ad ogni passaggio agglomerativo, i due cluster che si fondono sono quelli con il più piccolo incremento nella somma totale del quadrato delle distanze all'interno del cluster.

Nella figura 3.4 viene riportato il dendrogramma risultante alla clusterizzazione del test Eurofit.

### 3.4) Applicazione dell'analisi cluster nel test Eurofit



*Figura 3.4 : Albero di raggruppamento nell'Eurofit*

Si nota come effettivamente si potrebbero ottenere due grossi sottogruppi, forse legati tra loro da alcune caratteristiche in comune. Si effettua il taglio dell'albero in due, come indicato dai due rettangoli. A seguito del taglio, con dei boxplot si va ad analizzare le variabili precedentemente rilevate in funzione dei due nuovi gruppi generati.

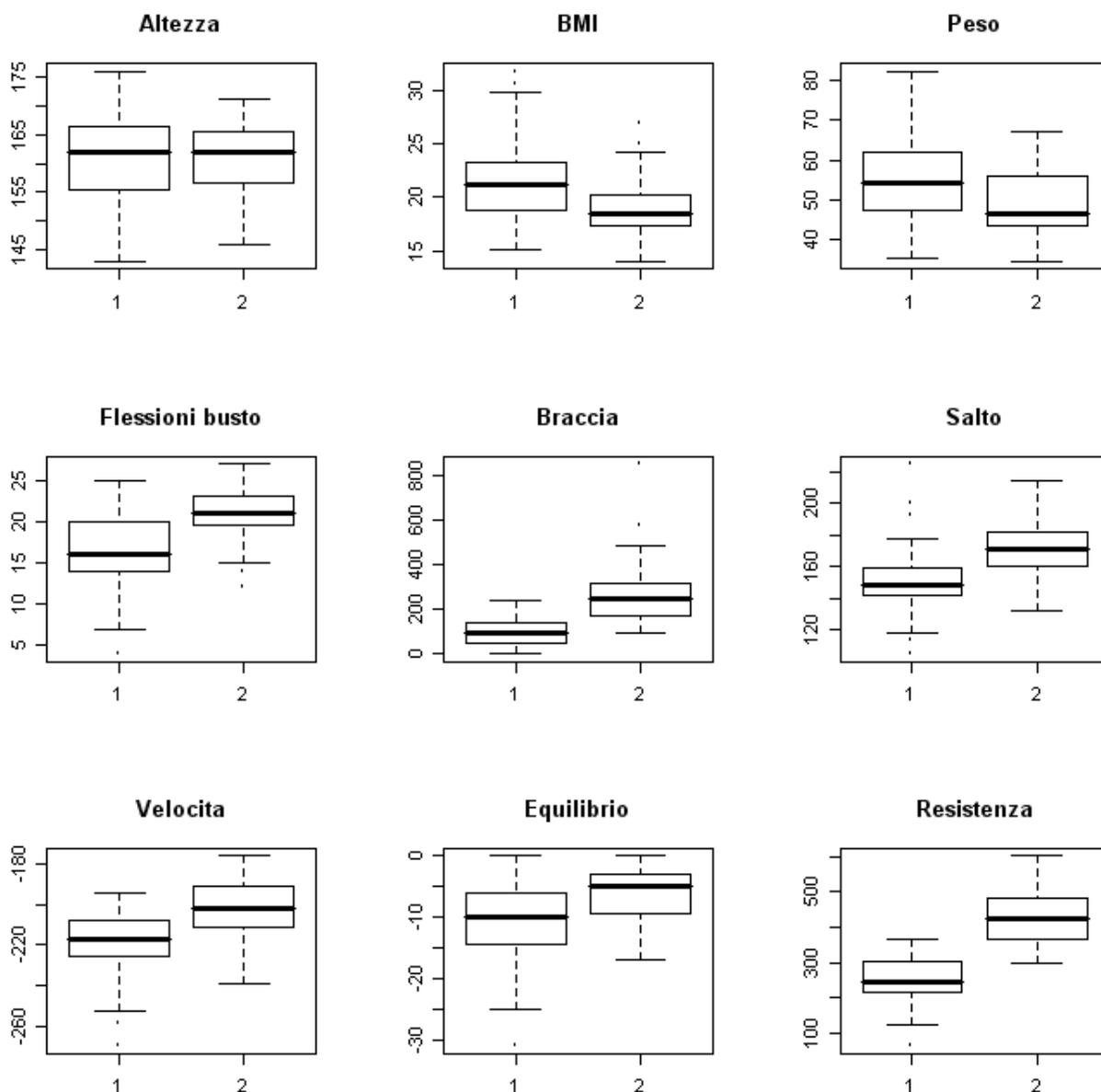


Figura 3.5 : Boxplot delle variabili nei due gruppi formati

I primi tre boxplot nella prima riga identificano le misure antropometriche dei due gruppi. A parità d'altezza, il primo gruppo pesa di più, e di conseguenza anche il BMI risulta più elevato. Si nota chiaramente come il primo gruppo, quello costituito da studenti più pesanti, sia anche quello che ottiene peggiori risultati. Negli altri sei boxplot infatti il primo gruppo è più basso del secondo, in ogni campo. I due gruppi formati con l'albero confermano quanto analizzato precedentemente con le correlazioni.

### 3.5) Applicazione dell'ACP nel questionario Psdq

Verrà usata la tecnica dell'ACP nel questionario Psdq, alla ricerca di nuovi aspetti sui legami nelle variabili presenti. Si è ritenuto opportuno utilizzare le prime due componenti principali per spiegare il 61% della varianza totale.

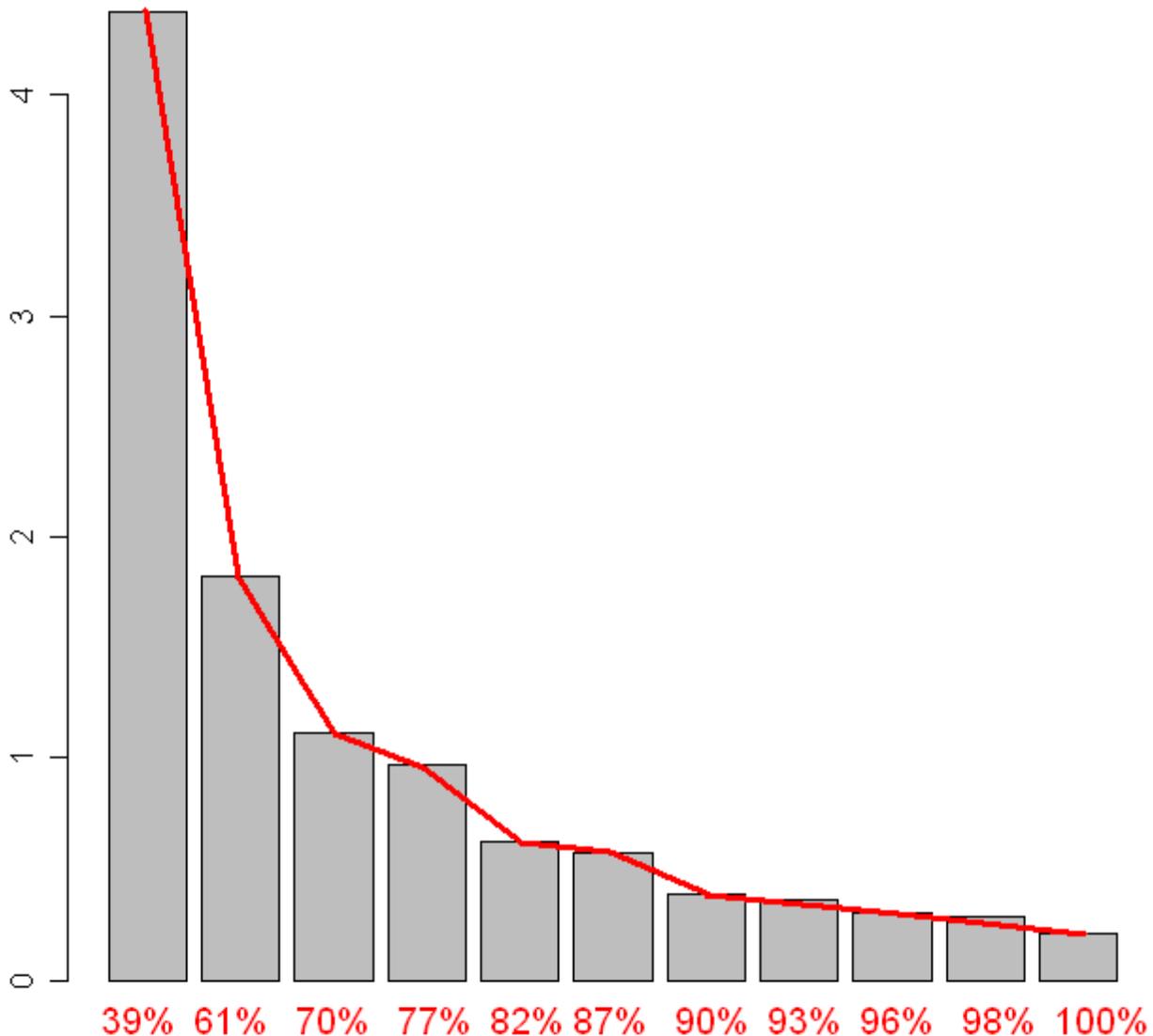


Figura 3.6 : Gli Autovalori, il punto di gomito e la varianza spiegata

Anche in questo caso ACP offre risultati interessanti. Si riesce a identificare i due assi (Figura 3.7). Quello orizzontale, sembra identificare **la qualità del proprio aspetto esteriore**, quello verticale invece il **la qualità del proprio fisico**. Verso sinistra infatti notiamo la valutazione del proprio fisico, del proprio aspetto e l'autostima, che viaggiano tutte nella stessa direzione. Verso l'alto abbiamo quattro concezioni che valutano tutte potenzialità fisiche. La variabile che sembra andare in direzione opposta alla direzione del "piacersi fisicamente" è la concezione di obesità, e questo rispecchia quanto già dapprima rilevato. Le variabili sono state rinominate come riportato nella Tabella 3.2.

<i>Variabile originale</i>	<i>Variabile rinominata</i>
Salute	sal
Coordinazione	coo
Attività fisica	att
Obesità	obe
Abilità fisica	abi
Fisico	fis
Aspetto fisico	asp
Forza	for
Resistenza	res
Stima	sti
Flessibilità	fle

Tabella 3.2 : Rinominazione delle variabili relative al questionario Psdq

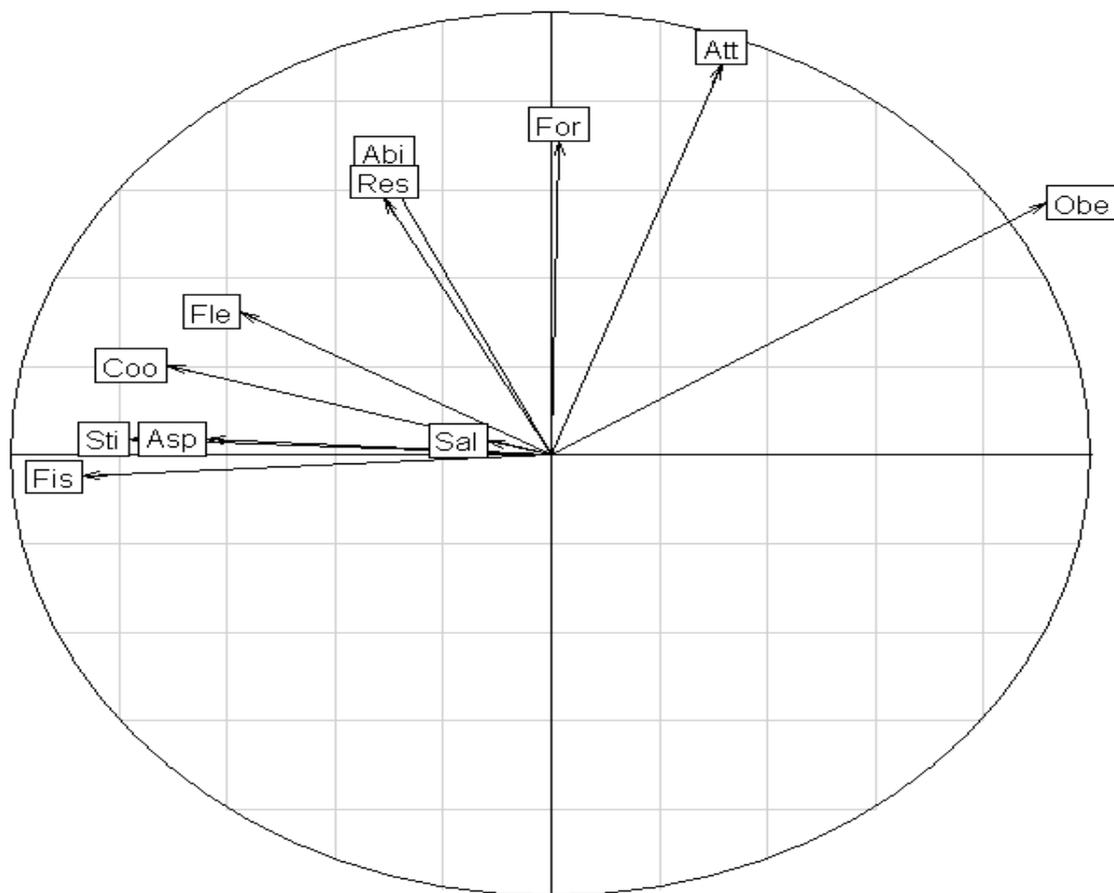


Figura 3.7 : Cerchio delle correlazioni nel questionario Psdq

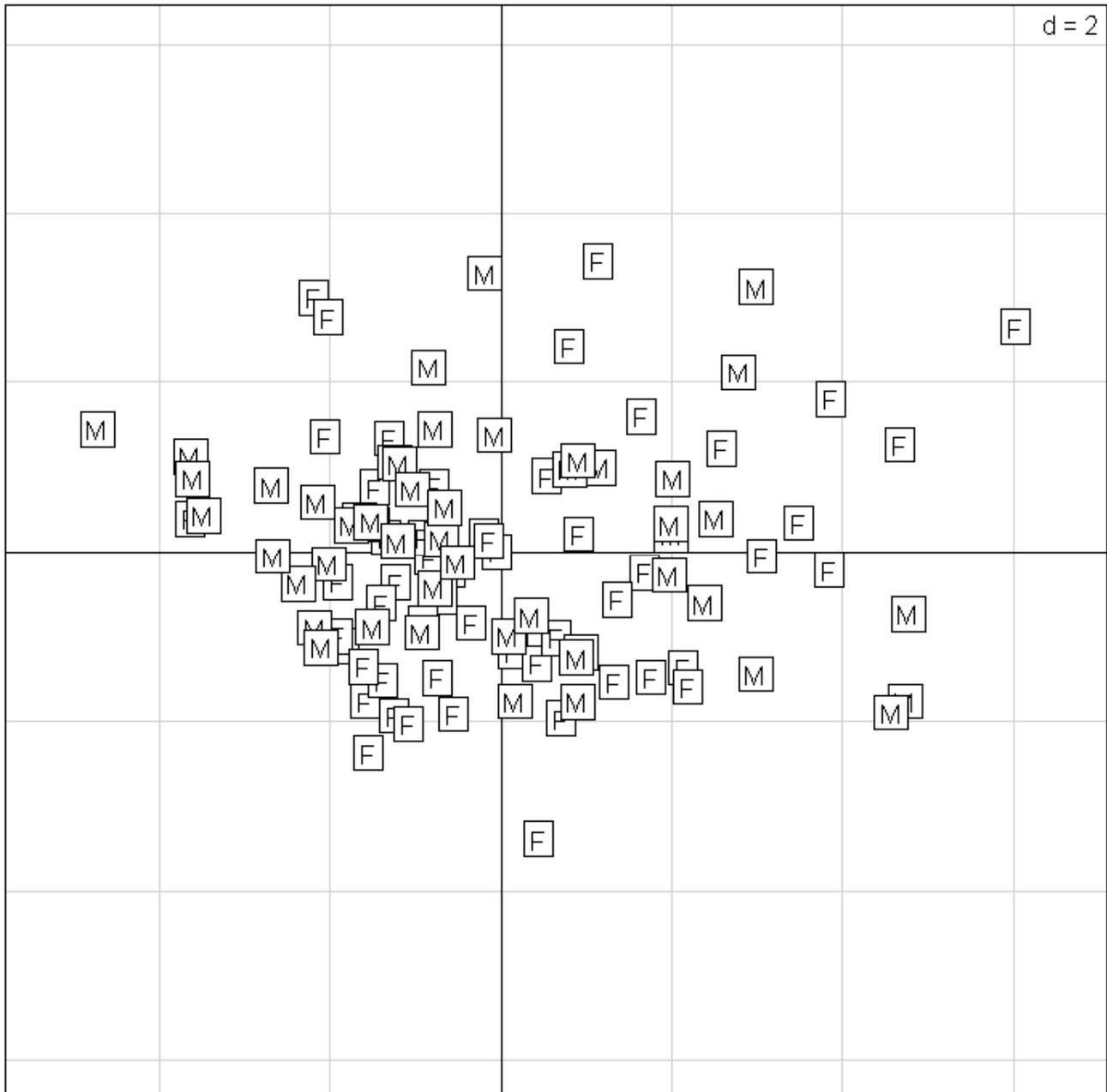


Figura 3.8 : Rappresentazione grafica dei punteggi fattoriali

Analizzando la figura 3.8 notiamo una concentrazione maggiore di studenti che tende a posizionarsi nella parte centrale sinistra del grafico, interpretabile col cerchio delle correlazioni come una zona dove i ragazzi tendono a piacersi.

Nella figura 3.9 è ripotato il dendrogramma alla ricerca di gruppi omogenei.

### 3.6) Applicazione dell'analisi cluster nel questionario Psdq

#### Dendrogramma

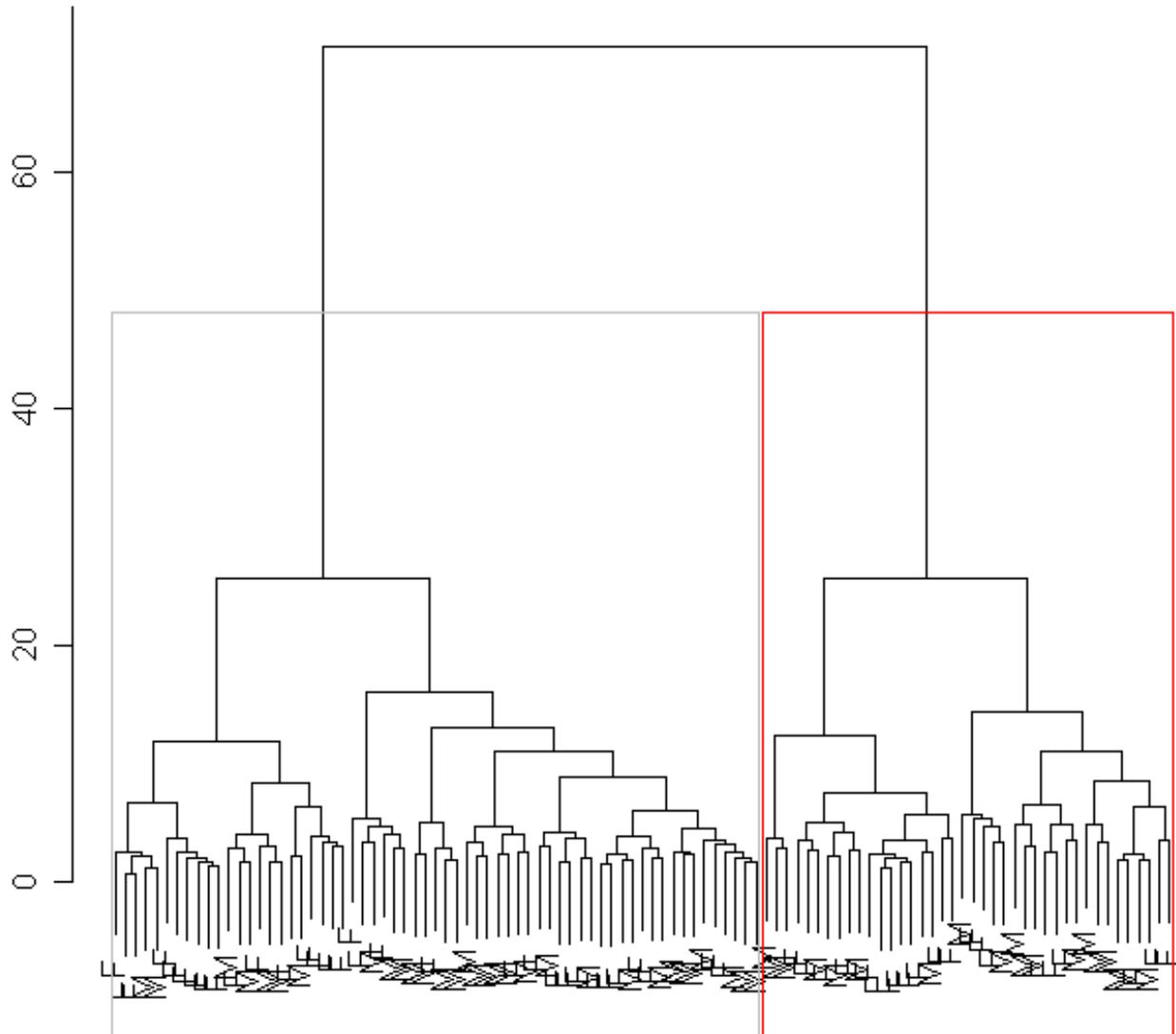


Figura 3.4 : Albero di raggruppamento nel Psdq

Anche in questo caso l'albero è stato diviso in due gruppi cercando di capire poi quale sia l'aspetto che ha portato alla divisione del campione nelle due nuove popolazioni. Precedentemente, nell'analisi del test Eurofit, si era notato come fosse il Bmi che maggiormente andava a differenziare i due gruppi.

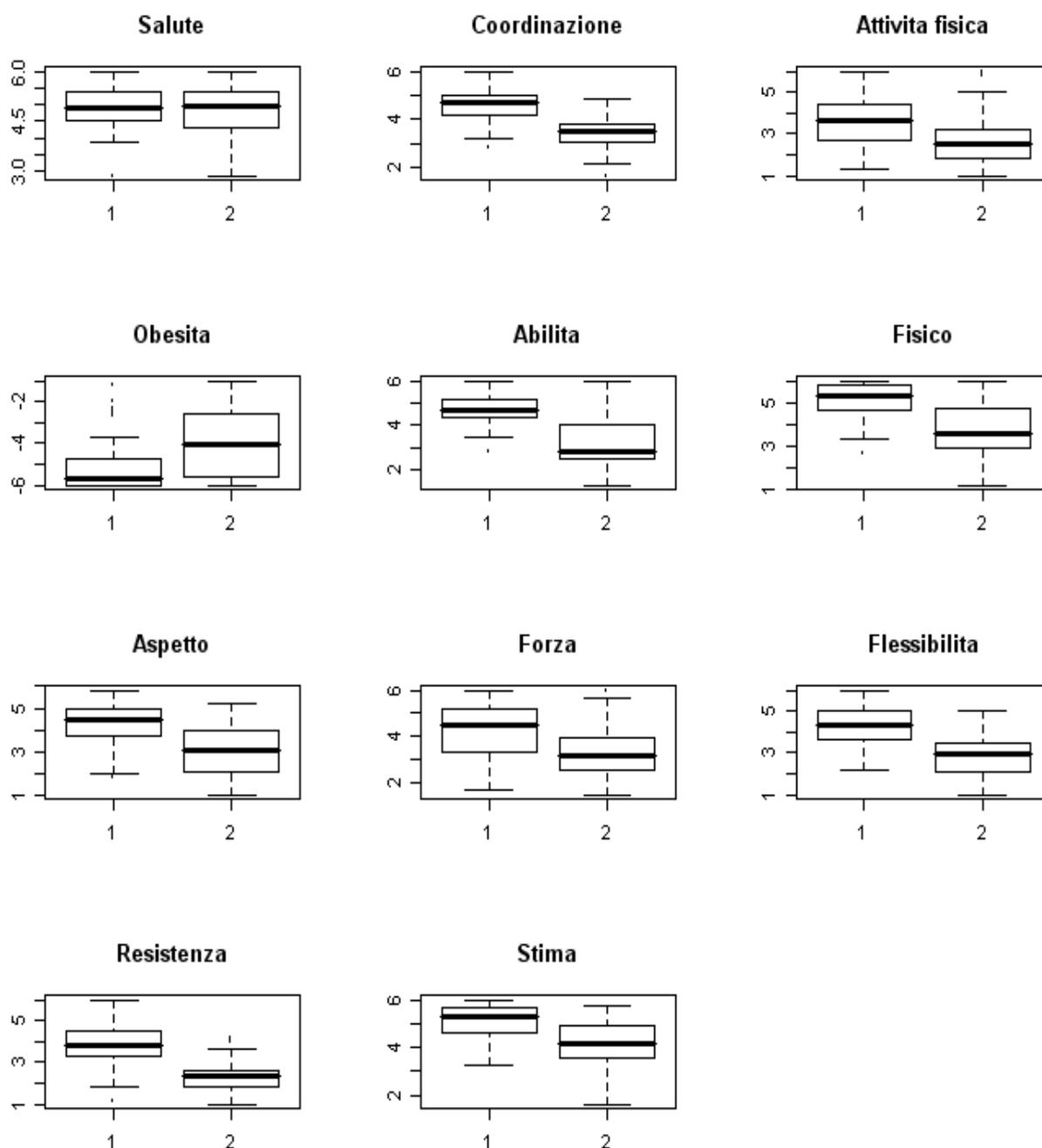


Figura 3.10 : Boxplot relativi ai due gruppi formati

I risultati confermano quanto prima individuato con il test Eurofit. Si nota chiaramente come, a parità di condizione di salute, l'unica variabile che lavora al contrario rispetto alle altre è il sentirsi obeso. Il primo gruppo, caratterizzato da persone che si vedono magre e che, vista la correlazione forte determinata precedentemente, lo sono davvero, si valutano estremamente meglio del gruppo caratterizzato da persone che si sentono obese. Soprattutto per quanto riguarda l'aspetto fisico e la propria autostima, che non erano correlati con le potenzialità fisiche, anche con

questa tecnica si nota chiaramente la differenza imposta dall'obesità.

### 3.7) L'ACP nella totalità delle variabili

Visto l'elevato numero di variabili, il numero di componenti principali aumenta rispetto ai casi precedenti. Nell'analisi condotta ci si limita all'uso di quattro fattori, arrivando al 56% di varianza spiegata.

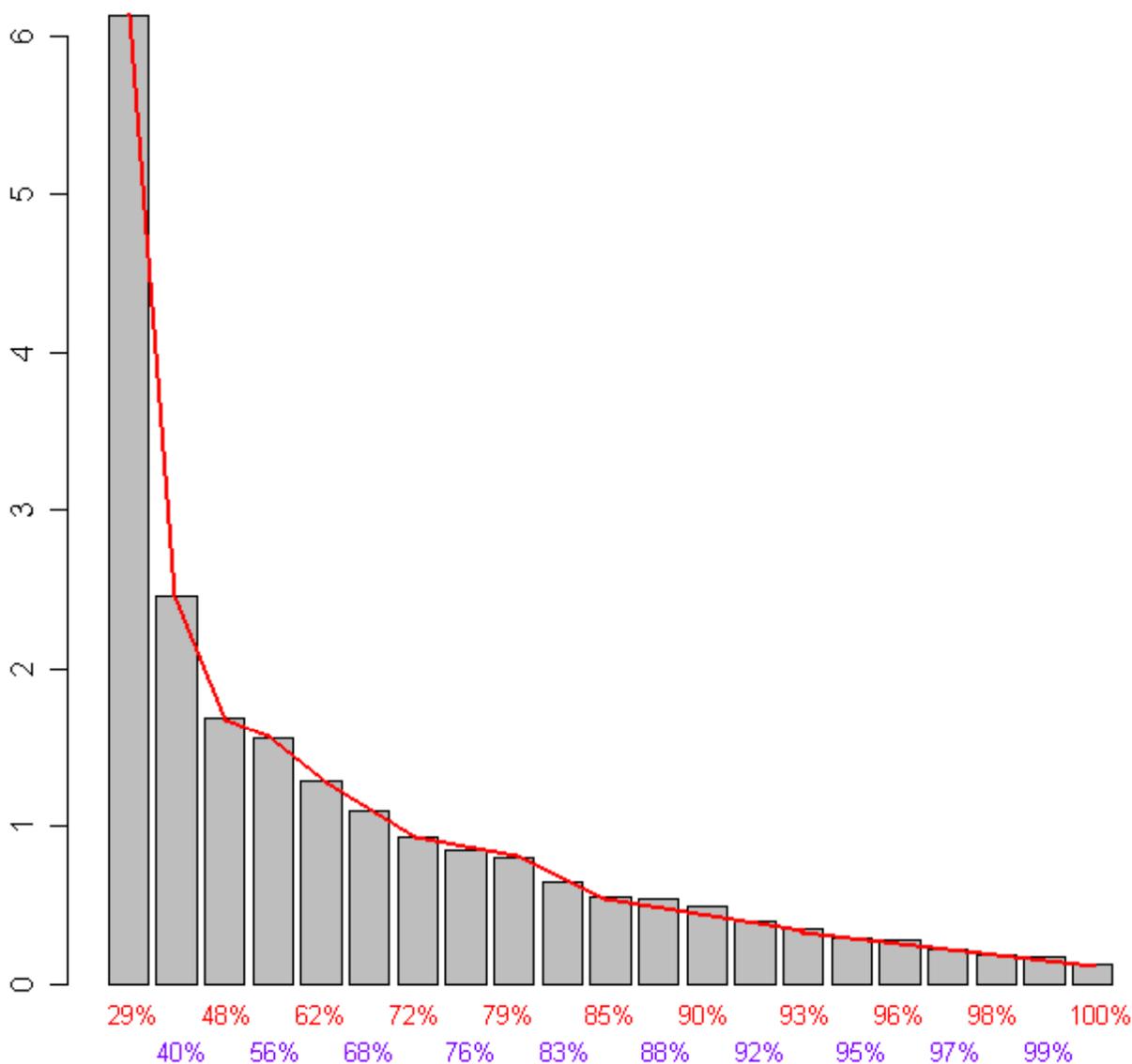


Figura 3.11 : Gli autovalori, il punto di gomito e la varianza spiegata

Nella figura 3.12 è riportato il cerchio delle correlazioni, e la relativa interpretazione. Le variabili seguono la stessa rinominazione precedentemente descritta, ma viene aggiunto un "P" maiuscolo come estensione delle variabili che arrivano dal questionario Psdq.

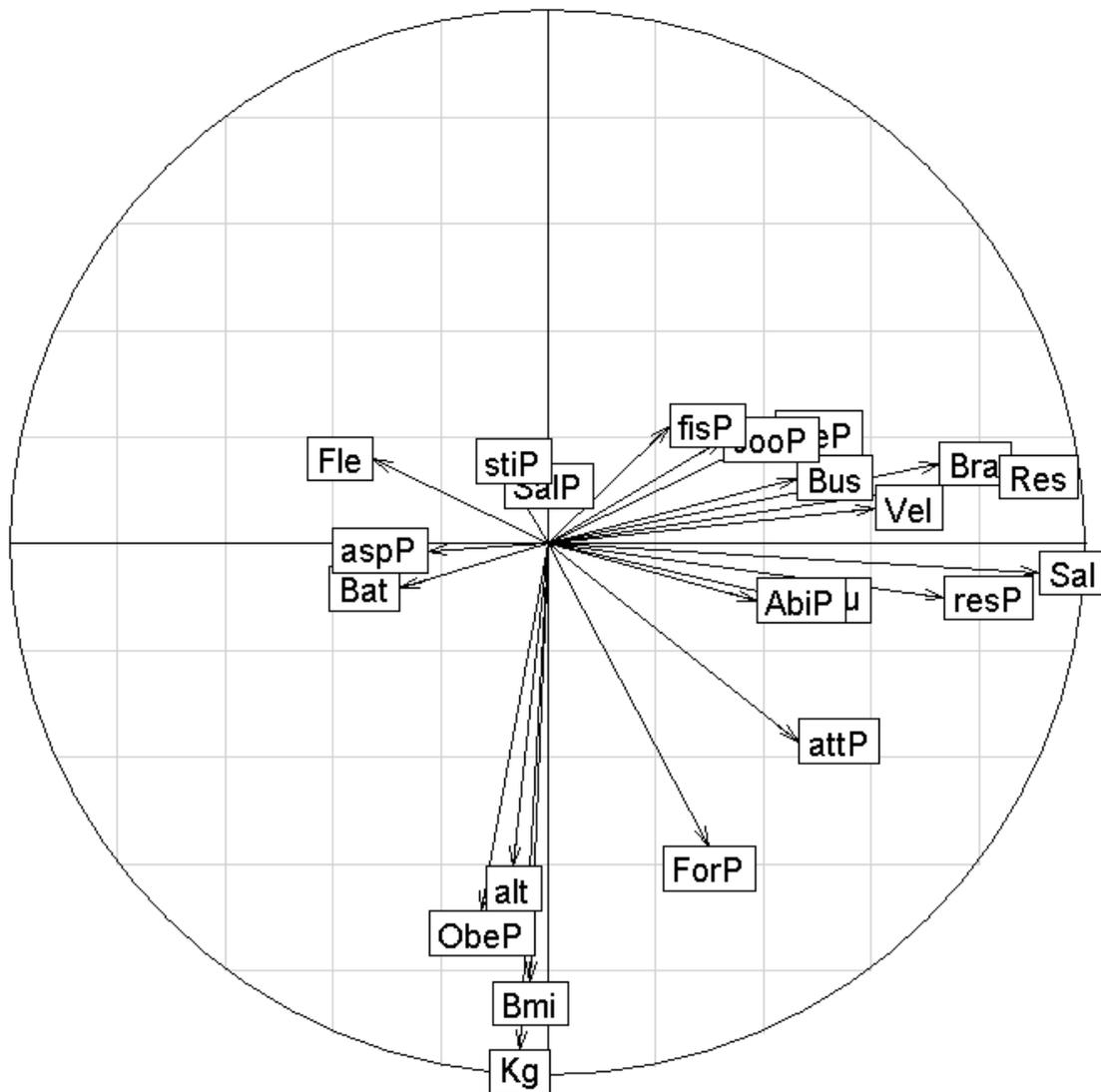


Figura 3.12 : Cerchio delle correlazioni del primo e il secondo fattore. sull'asse orrizontale la prestazione fisica, su quello verticale la robustezza

Questo primo cerchio studia le correlazioni tra il primo e il secondo fattore. L'asse verticale, denota come vi sia un legame tra le misure antropometriche e la concezione di forza. E' plausibile che un ragazzo che si vede grande e grosso tenda a credere di essere anche forte. Il primo asse quindi può prendere il nome di **robustezza**, mentre il secondo asse denota le **prestazioni fisiche**. Segue ora il cerchio realizzato utilizzando il secondo e il terzo fattore.

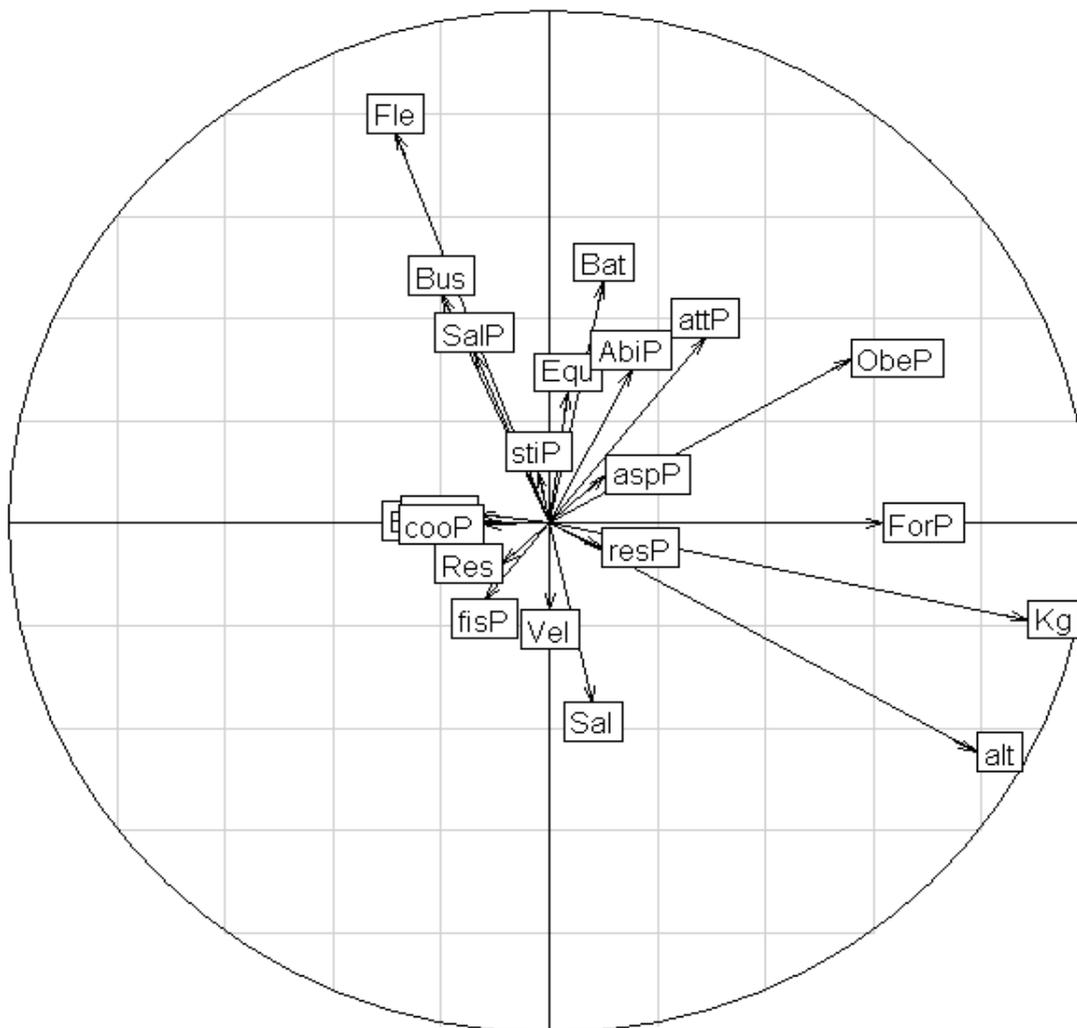


Figura 3.13 : Cerchio delle correlazioni tra il secondo e il terzo fattore. Si nota la predominante saturazione di Flessioni col terzo fattore(asse verticale).

Questo cerchio permette di notare un aspetto particolare delle variabili utilizzate. Mentre la maggior parte delle caratteristiche ruotano attorno al centro del cerchio, con una saturazione quindi limitata, l'asse verticale sembra venir correlato fortemente dalla variabile Fle , ovvero Flessioni del busto. Si conferma quindi la particolarità di questo esercizio ginnico rispetto agli altri della batteria Eurofit. Nell'asse orizzontale ritroviamo il concetto di **robustezza**. Per ultimo, il cerchio delle correlazioni prodotte col terzo e il quarto fattore.

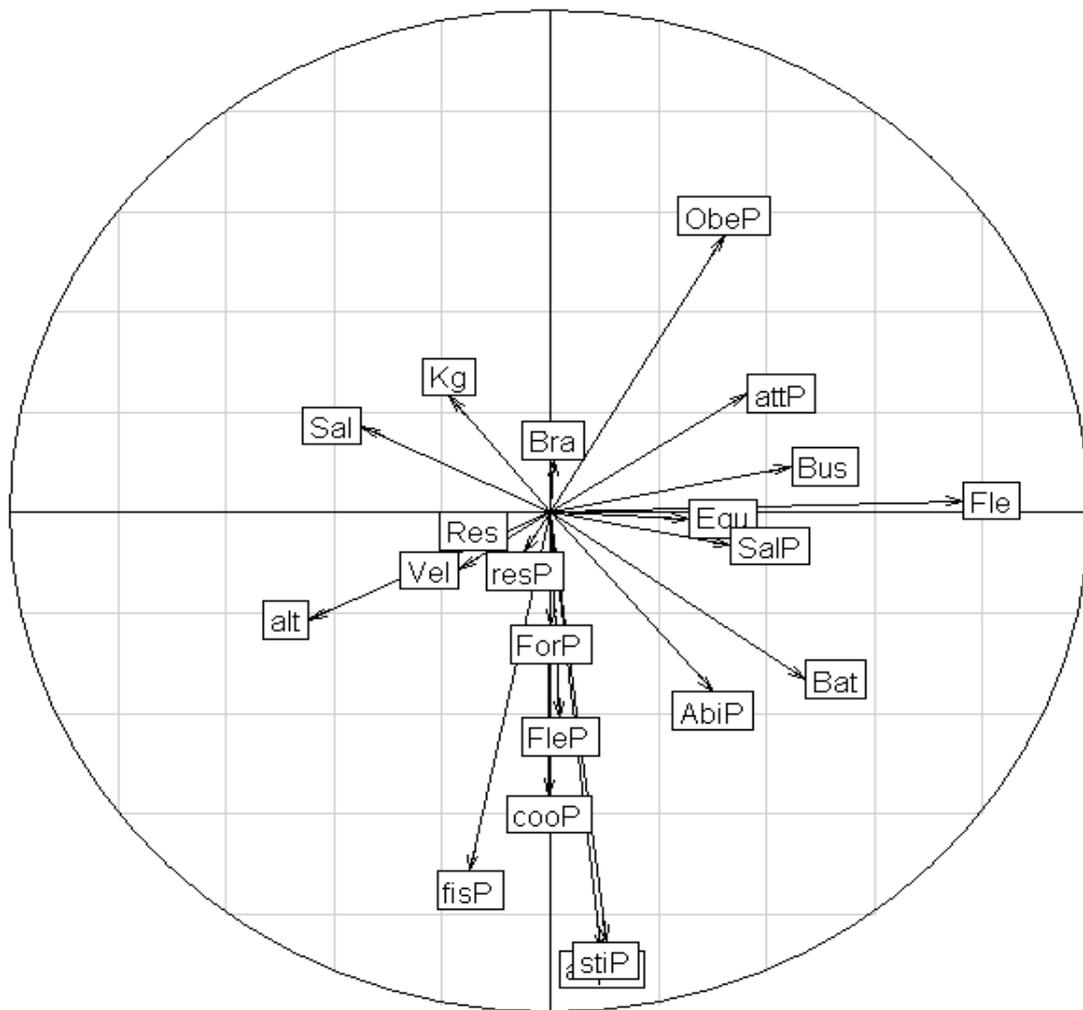


figura 3.14 : Cerchio delle correlazioni del terzo e il quarto fattore

Anche in questo caso Flessioni è la variabile con la maggiore saturazione con l'asse orizzontale, portando alle stesse conclusioni del caso precedente. L'asse verticale, quello del terzo fattore, sembra riassumere il concetto di **Concezione del sè**, infatti troviamo la maggior parte delle variabili provenienti dal questionario Psdq con una forte correlazione con l'asse. L'asse orizzontale invece sintetizza le **caratteristiche fisiche**, non legate alla concezione psicologica della propria persona. I cerchi ottenibili con le altre combinazioni portano alle stesse conclusioni, e quindi non vengono riportati.



## Conclusioni

L'analisi condotta porta a concludere che non ci sia una forte associazione tra il questionario Psdq e le variabili rilevate nel test Eurofit. Di conseguenza, o i test veramente non possono essere associati, in quanto non stanno rilevando la stessa cosa, oppure i ragazzi non riescono a valutarsi correttamente. I ragazzi riescono a percepire correttamente il proprio fisico solamente nel valutare il proprio stato di obesità e di resistenza fisica. Questo non avviene invece per la forza muscolare, la flessibilità e la coordinazione. Poche le differenze nei due generi, sia a livello psicologico che potenziale, evidenziate solo nella maggior resistenza muscolare per i maschi e nella miglior flessibilità per le femmine. A seguito dell'analisi emerge un risultato importante, cioè che l'autostima che questi ragazzi hanno non dipende da quanto riescono potenzialmente a fare, ma danno una grossa importanza al proprio aspetto fisico e al lato esteriore. Un aspetto molto importante gira attorno al concetto di obesità. L'obesità infatti porta deficit non indifferenti nelle capacità fisiche e soprattutto nel modo di vedersi e concepirsi. Sembra inoltre esserci un effetto di auto limitazione nei ragazzi che si vedono obesi, in quanto sono state individuate dipendenze tra il sentirsi obesi e il valutarsi fisicamente poco energici che non son presenti poi a valutare i risultati oggettivi. La letteratura sull'argomento evidenzia che quest'effetto secondario esiste ed è già stato studiato da psicologi dello sport. Vengono citate alcune affermazioni del dottor Guicciardi, docente di psicologia dello sport presso la Facoltà di Scienze della Formazione di Cagliari :

*"Non ce la faccio, non ce la farò mai". Quante volte capita di sentirsi inadeguati davanti alla prospettiva di uno sforzo fisico? Da qualche anno questa sensazione ha un nome: scarsa "autoefficacia".*

*"Il concetto di self-efficacy, che noi traduciamo con autoefficacia, è nato nel 1977", spiega il docente, "e può essere definito come la fiducia che una persona ripone nelle proprie capacità di eseguire con successo un compito". L'autoefficacia diventa una variabile molto importante, perché le convinzioni di efficacia influenzano i modi di pensare, la scelta delle azioni da intraprendere, la perseveranza di fronte agli ostacoli, il grado di coinvolgimento e d'impegno nel compito, la capacità di resistere alle avversità .*

Quanto sostenuto dal docente di psicologia è confermato nell'analisi, dove si dimostra come condizioni psicologiche possono andar a influenzare le prestazioni fisiche. Il freno maggiore lo crea l'obesità, che genera a livello psicologico più limiti di quanto ne crei a livello fisico. Un fisico in sovrappeso quindi non solo porta malus prestazionali nello sport, ma anche nelle condizioni di vita.



## APPENDICE

### **Importazione dei dati da file excel:**

salvare il documento excel come un file csv, separatore selezionare ";" e salvarlo nella directory di lavoro di R.

In R eseguire

```
nomedataset <- read.csv("nomefile.csv",header=T,sep=";")
```

### **Comandi per l'analisi in componenti principali**

si carica la libreria ade4 con `library(ade4)` dopo averla installata con "installa pacchetti"

si estraggono le componenti principali

```
estrazione <- dudi.pca(dataset)
```

si genera il cerchio delle correlazioni

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

oppure prima si ruotano i fattori, e poi si genera il cerchio delle correlazioni

```
rotazione <- promax(as.matrix(rotazione$co))
```

```
s.corcircle(rotazione$loadings[,c(x,y)])
```

dove x,y indicano su che fattori proiettare il cerchio delle correlazioni

### **Comandi per l'analisi cluster**

Si carica la relativa libreria

```
library(cluster)
```

Si calcolano le distanze col metodo specificato in method

```
dista = dist(dataset,method="euclidean")
```

si genera l'albero formando i gruppi col metodo specificato

```
h1 = hclust(dista,method="ward")
```

si taglia l'albero in k gruppi

```
patiz1 = cutree(h1,k=2)
```



## Riferimenti Bibliografici

Fabbris L (1990), "Analisi esplorative di dati multidimensionali", Cleup Editore.

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

A. Azzalini B. Scarpa (2004), "Analisi dei dati e Data Mining", Springer.