

ANALISI DISCRIMINANTE E MANOVA

In questo documento presentiamo alcune opzioni analitiche della procedura di analisi discriminante e della procedura GLM di SPSS che non sono state incluse nel testo pubblicato. Si tratta di opzioni che, pur non essendo utilizzate correntemente per effettuare l'analisi fattoriale, consentono di completare lo spettro delle possibilità di analisi offerte da SPSS. In particolare, il documento completa la descrizione della procedura di analisi dei cluster con in seguenti paragrafi:

1. Analisi Discriminante:

1.1. Analisi Discriminante con il linguaggio SINTASSI

1.2. Statistiche descrittive

1.3. Coefficienti non standardizzati

1.4. Probabilità a priori e coefficienti delle funzioni di classificazione

1.5. Rappresentazioni grafiche: Mappa territoriale e grafici a dispersione per gruppi separati

1.6. Analisi discriminante con il metodo stepwise

1.7. Analisi discriminante con il metodo gerarchico

2. MANOVA:

2.1. MANOVA con il linguaggio SINTASSI

2.2. L'Analisi Multivariata della Covarianza (MANCOVA)

1. Analisi Discriminante:

1.1. Analisi Discriminante con il linguaggio SINTASSI

Di seguito presentiamo i programmi SINTASSI per le analisi descritte nel testo.

```
**** Analisi discriminante: modello base.
```

```
DISCRIMINANT  
  /GROUPS=clu4_1(1 4)  
  /VARIABLES=n e o a c  
  /ANALYSIS ALL  
  /PRIORS EQUAL  
  /STATISTICS=MEAN STDDEV UNIVF BOXM  
              COEFF RAW TABLE CROSSVALID  
  /PLOT=COMBINED CASES MAP  
  /CLASSIFY=NONMISSING POOLED .
```

Illustriamo brevemente i comandi specificati nella procedura DISCRIMINANT.

Il comando `/GROUPS=clu4_1(1 4)` serve a definire la variabile dipendente ovvero la variabile che specifica i raggruppamenti pre-definiti di soggetti. L'utente deve specificare i valori minimo e massimo che vengono utilizzati per differenziare i gruppi (nel nostro caso abbiamo quattro gruppi, che hanno etichette progressive da 1 a 4). Non è possibile ovviamente specificare più di una variabile dipendente, come del resto risulta chiaro dalla procedura eseguita tramite menù (vedi pag. 174 del testo).

Il comando `/VARIABLES=n e o a c` serve invece a definire le variabili indipendenti che vengono inserite per rendere ragione della appartenenza dei soggetti ai diversi gruppi definiti dalla variabile dipendente. Le variabili devono essere a intervalli equivalenti o dicotomiche (vedi Barbaranelli, 2003).

Il comando `/ANALYSIS ALL` può essere utilizzato per definire diverse analisi discriminanti sugli stessi dati, ma analizzando insieme diversi di variabili. Nel nostro caso vengono analizzate tutte le variabili indipendenti specificate col comando precedente.

Il comando `/PRIORS EQUAL` serve per definire le probabilità a priori di appartenenza ai gruppi, utilizzate nella fase di classificazione. L'opzione di default è che un soggetto abbia la stessa probabilità di essere classificato in ciascun gruppo (ed è quella specificata nel nostro caso).

Ovviamente l'utente può specificare valori differenti, se questo si rende necessario (per l'opportunità e le conseguenze di questa operazione a Barbaranelli, 2003). La somma dei valori di probabilità a priori forniti dall'utente deve essere uguale a 1 e l'utente deve fornire tanti valori quanti sono i gruppi. Se l'utente specifica un valore pari a 0 per un gruppo, nessun soggetto verrà classificato in quel gruppo. Se l'utente specifica "SIZE", allora la probabilità a priori verrà derivata dalla numerosità osservata di ciascun gruppo.

Il comando `/STATISTICS` serve a richiedere una serie di statistiche supplementari che possono arricchire l'output oltre quanto viene fornito di default. Di seguito forniamo una descrizione di quelle specificate nel nostro esempio. `MEAN` e `STDDEV` servono per richiedere le medie e le deviazioni standard delle variabili indipendenti nei gruppi. `UNIVF` serve per richiedere i test univariati dell'analisi della varianza che esaminano la significatività della differenza delle medie delle variabili indipendenti tra i differenti gruppi. `BOXM` serve per richiedere il test M di Box per la verifica dell'assunzione dell'omogeneità delle matrici di varianza/covarianza. `COEFF` serve per richiedere in output i coefficienti delle funzioni di classificazione: anche se la procedura non utilizza questi coefficienti per classificare i casi utilizzati nell'analisi, essi possono essere utilizzati per classificare altri soggetti che appartengono a campioni differenti. `RAW` consente di ottenere i coefficienti non standardizzati delle funzioni discriminanti (vedi Barbaranelli, 2003 per una definizione di tali coefficienti). `TABLE` consente di visualizzare i risultati del processo di classificazione dei dati: nel caso in cui l'analisi fosse stata condotta selezionando un gruppo di soggetti dal file, la tabella conterrà due sezioni separate, una per i soggetti utilizzati nell'analisi e una per i soggetti non selezionati. `CROSSVALID` consente di effettuare il processo di classificazione tramite la procedura della "classificazione autoesclusiva" (si veda pag. 179 del testo, ma anche Barbaranelli, 2003). Infine, specificando "ALL", verranno riportate in output tutte le

statistiche disponibili, quindi sia quelle di default, sia quelle descritte ora, sia altre che non abbiamo descritto e che in parte vengono descritte nel testo poiché ottenibili da menù.

Il comando `/PLOT= COMBINED CASES MAP` serve ad ottenere alcune rappresentazioni grafiche che possono aiutare il ricercatore nell'interpretazione dei risultati. In particolare, specificando `COMBINED` si ottiene la rappresentazione grafica dei casi che appartengono a tutti i gruppi analizzati, utilizzando le prime funzioni discriminanti come coordinate (si tratta del diagramma nella figura 4.9 a pag. 192 del testo). Specificando invece `SEPARATE` si ottiene una rappresentazione grafica separata per ognuno dei gruppi considerati in analisi (come quella che viene riportata nella figura 2 di questo documento). Specificando `CASES` si ottengono in output le statistiche della classificazione che vengono riportate nella 4.9 a pag. 194 del testo: rimandiamo dunque al testo per la descrizione di tali indici. Specificando “`CASES(n)`” le statistiche di classificazione vengono riportate solo per i primi `n` soggetti, altrimenti vengono riportate per tutti i casi. Specificando `MAP` si ottiene in output la mappa territoriale, una rappresentazione grafica che viene descritta nel paragrafo 1.5 di questo documento.

Il comando `/CLASSIFY=NONMISSING POOLED` serve per gestire i casi nel processo di classificazione. In particolare, l'opzione `NONMISSING` specifica che verranno classificati solo i casi che non presentano valori mancanti, mentre l'opzione `POOLED` specifica che nelle procedure di classificazione verrà utilizzata la matrice “*pooled within sample*” (o di covarianza entro i gruppi). Opzioni ulteriori servono per: a) classificare solo i casi non selezionati (“`UNSELECTED`”); b) utilizzare le matrici di covarianze di ogni gruppo separato (“`SEPARATE`”); c) classificare solo i casi non classificati (“`UNCLASSIFIED`”); d) sostituire i valori mancanti delle variabili indipendenti con la media (“`MEANSUB`”).

1.2. Statistiche descrittive

La tabella 1 presenta le statistiche descrittive basilari (media, deviazione standard, numero di casi) per i 4 gruppi e per il campione totale nell'esempio descritto alle pagine 180-196 del testo. La

tabella 2 presenta la tabella del test della differenza delle medie (ANOVA) per le 5 variabili nei 4 gruppi. Entrambe le tabelle sono ottenibili dalla finestra di dialogo Statistiche (vedi figura 4.4 pag. 175). Le medie risultano significativamente diverse nei 4 gruppi, a riprova del fatto che i gruppi contengono soggetti con profili piuttosto diversificati nelle 5 variabili.

Tabella 1. Statistiche descrittive dei gruppi

Statistiche di gruppo					
CLU4_1 Ward Method		Media	Deviazione std.	Validi (listwise)	
				Non pesati	Pesati
1	N	1,9524	,41682	74	74,000
	E	2,6208	,30650	74	74,000
	O	2,8916	,23563	74	74,000
	A	2,8431	,33109	74	74,000
	C	2,3431	,31486	74	74,000
2	N	2,2977	,35194	134	134,000
	E	2,1144	,35789	134	134,000
	O	2,5127	,30629	134	134,000
	A	2,3217	,41886	134	134,000
	C	2,3528	,44908	134	134,000
3	N	1,8389	,38202	156	156,000
	E	2,2675	,31438	156	156,000
	O	2,4169	,35494	156	156,000
	A	2,6677	,29924	156	156,000
	C	2,8974	,31999	156	156,000
4	N	1,4491	,38779	54	54,000
	E	2,7207	,30878	54	54,000
	O	2,9055	,27217	54	54,000
	A	2,8045	,48019	54	54,000
	C	3,2922	,28167	54	54,000
Totali	N	1,9557	,46820	418	418,000
	E	2,3395	,39652	418	418,000
	O	2,5948	,37151	418	418,000
	A	2,6055	,42451	418	418,000
	C	2,6757	,50056	418	418,000

Tabella 2. Test di uguaglianza delle medie tra I gruppi

Test di uguaglianza delle medie di gruppo

	Lambda di Wilks	F	df1	df2	Sig.
N	,654	73,135	3	414	,000
E	,675	66,397	3	414	,000
O	,695	60,636	3	414	,000
A	,764	42,555	3	414	,000
C	,518	128,347	3	414	,000

Le tabelle 3a e 3b presentano rispettivamente le matrici di correlazione e covarianze *entro i gruppi* e quelle di covarianza calcolate su ogni gruppo separato e complessivamente su tutti i soggetti. Le prime due sono le cosiddette matrici *Pooled Within Group* e vengono calcolate effettuando una media ponderata delle matrici di correlazione e di covarianza di ciascun gruppo.

Tabella 3a. Matrici di covarianza correlazioni entro i gruppi (*pooled within*)

Matrici globali entro gruppi^a

		N	E	O	A	C
Covarianza	N	,144	-,017	,017	,018	,007
	E	-,017	,107	,009	,005	,015
	O	,017	,009	,097	,016	-,001
	A	,018	,005	,016	,139	-,013
	C	,007	,015	-,001	-,013	,131
Correlazione	N	1,000	-,140	,145	,126	,052
	E	-,140	1,000	,093	,042	,125
	O	,145	,093	1,000	,135	-,010
	A	,126	,042	,135	1,000	-,097
	C	,052	,125	-,010	-,097	1,000

a. La matrice di covarianza ha 414 gradi di libertà.

Tabella 3b. Matrici di covarianza sui gruppi separati

Matrici di covarianza^a

CLU4_1 Ward Method		N	E	O	A	C
1	N	,174	,030	,020	-,035	-,018
	E	,030	,094	,005	-,011	,025
	O	,020	,005	,056	,014	-,003
	A	-,035	-,011	,014	,110	-,008
	C	-,018	,025	-,003	-,008	,099
2	N	,124	-,044	,011	,021	,003
	E	-,044	,128	-,008	-,006	,024
	O	,011	-,008	,094	,012	,006
	A	,021	-,006	,012	,175	-,047
	C	,003	,024	,006	-,047	,202
3	N	,146	-,025	,031	,035	,011
	E	-,025	,099	,022	,015	,006
	O	,031	,022	,126	,021	,001
	A	,035	,015	,021	,090	-,004
	C	,011	,006	,001	-,004	,102
4	N	,150	,006	-,012	,032	,041
	E	,006	,095	,024	,027	,003
	O	-,012	,024	,074	,010	-,022
	A	,032	,027	,010	,231	,039
	C	,041	,003	-,022	,039	,079
Totali	N	,219	-,064	-,005	-,029	-,078
	E	-,064	,157	,050	,046	,046
	O	-,005	,050	,138	,039	-6,22E-05
	A	-,029	,046	,039	,180	,023
	C	-,078	,046	-6,22E-05	,023	,251

a. La matrice di covarianza globale ha 417 gradi di libertà.

1.3. Coefficienti non standardizzati

La tabella 4 presenta i coefficienti non standardizzati delle funzioni discriminanti. Questi coefficienti possono essere utilizzati dal ricercatore per calcolare eventualmente i punteggi nelle funzioni discriminanti di nuovi soggetti non inclusi nel campione in analisi.

Tabella 4. Coefficienti non standardizzati delle funzioni discriminanti

Coefficienti delle funzioni discriminanti canoniche

	Funzione		
	1	2	3
N	-1,642	,064	,189
E	,529	1,642	-,220
O	,711	1,987	1,855
A	1,173	,210	-2,118
C	1,866	-1,527	,943
(Costante)	-7,920	-5,582	-1,674

Coefficienti non standardizzati

1.4. Probabilità a priori e coefficienti delle funzioni di classificazione

La tabella 5 fornisce le probabilità a priori utilizzata nella fase di classificazione. Ovviamente, cambiando l'opzione di default nell'opzione che calcola la probabilità a priori dalla dimensione effettiva dei gruppi, cambiano anche i valori nella tabella. La tabella 6 riporta i coefficienti delle funzioni di classificazione. Si tratta di coefficienti che non vengono utilizzati per classificare i soggetti presenti nel campione ma che possono essere invece utilizzati per classificare nuovi soggetti per i quali non si dispone dell'informazione di appartenenza alla variabile di raggruppamento. I nuovi soggetti vengono attribuiti al gruppo rispetto al quale presentano il punteggio più elevato in una delle 4 funzioni di classificazione.

Tabella 5. Probabilità a priori

Probabilità a priori per gruppi			
CLU4_1 Ward Method	A priori	Casi usati nell'analisi	
		Non pesati	Pesati
1	,250	74	74,000
2	,250	134	134,000
3	,250	156	156,000
4	,250	54	54,000
Totali	1,000	418	418,000

Tabella 6. Coefficienti delle funzioni di classificazione

Coefficienti della funzione di classificazione				
	CLU4_1 Ward Method			
	1	2	3	4
N	10,282	13,074	9,336	6,073
E	20,940	17,144	17,222	19,999
O	23,459	19,992	19,143	24,382
A	17,354	13,763	17,274	18,274
C	16,919	16,880	21,586	24,615
(Costante)	-117,272	-95,484	-106,943	-134,555

Funzioni discriminanti lineari di Fisher

1.5. Rappresentazioni grafiche: Mappa territoriale e grafici a dispersione per gruppi separati

La figura 1 riporta la mappa territoriale (*territorial map*): si tratta di un grafico che rappresenta i centroidi dei gruppi e i confini utilizzati per classificare i casi nei gruppi considerando le prime 2 funzioni discriminanti: mentre i numeri da 1 a 4 indicano quale è il confine dei gruppi che hanno l'omonima numerazione, l'asterisco "*" all'interno dell'area delimitata da uno stesso confine rappresenta il centroide del gruppo.

La figura 2 presenta i grafici a dispersione costruiti considerando le prime due funzioni discriminanti. Mentre nel testo abbiamo descritto il grafico dei *gruppi accorpati* nel quale i 4 gruppi

analizzati vengono rappresentati simultaneamente, nella figura 2 la rappresentazione è effettuata separatamente per ciascun gruppo. Come per il grafico dei gruppi accorpati, nella figura 2 i casi sono rappresentati in uno spazio bidimensionale individuato dalle prime due funzioni discriminanti, ogni caso è indicato da un rettangolino e i centroidi sono indicati da rettangoli più grandi.

Figura 1. Mappa territoriale

Mapa territoriale (Assumendo che tutte le funzioni eccetto le prime due siano uguali a zero)

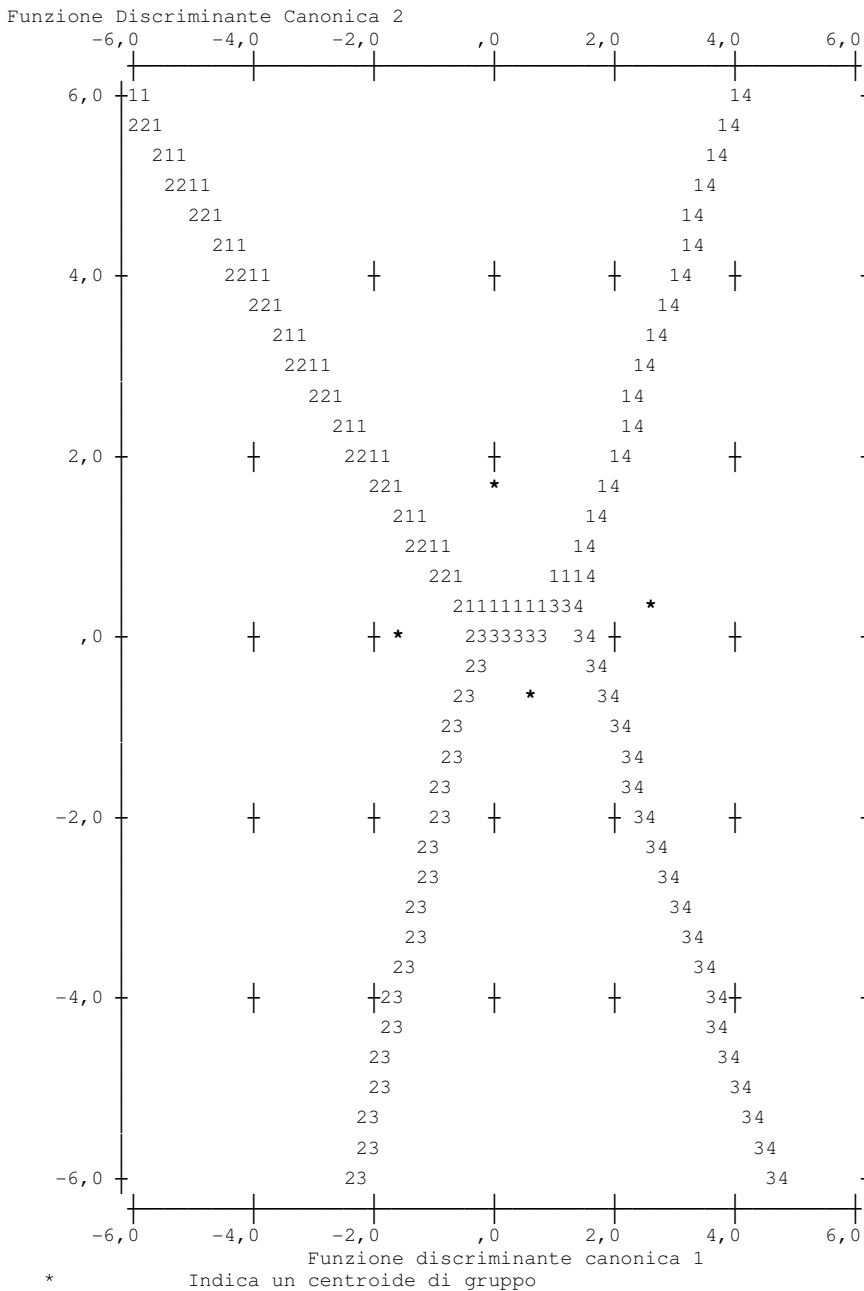
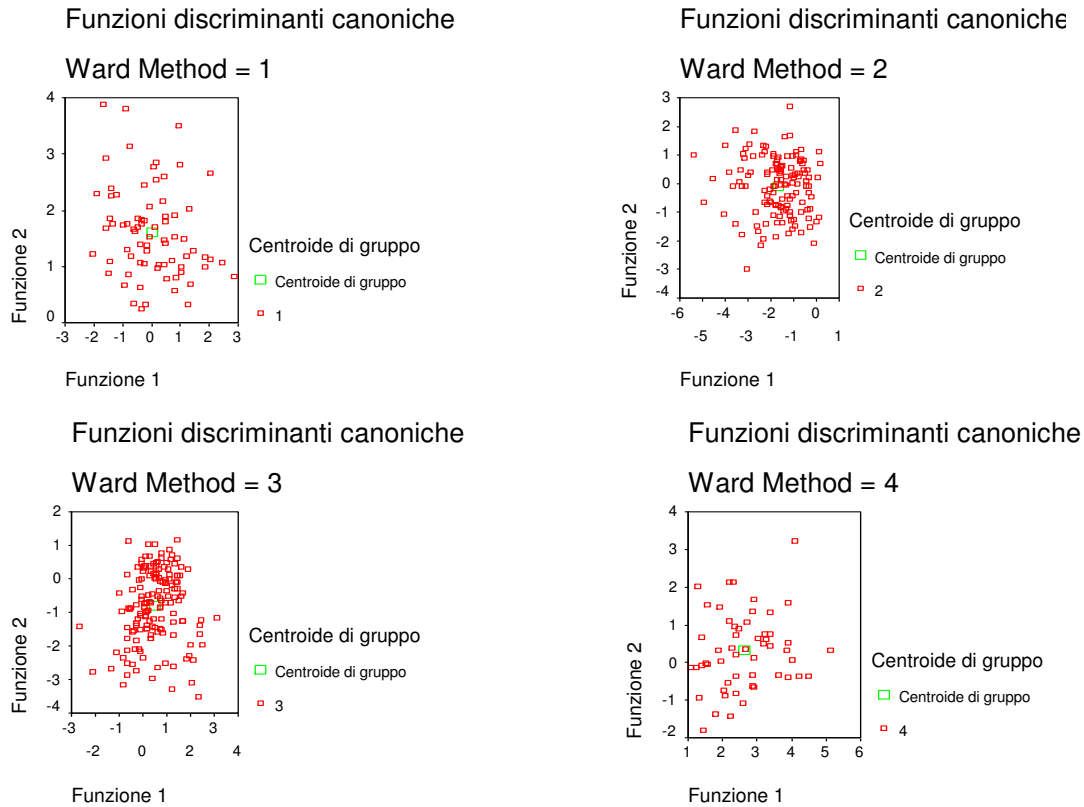


Figura 2. Diagramma di dispersione dei gruppi separati



Sia per i diagrammi di dispersione sia per la mappa territoriale, non è possibile utilizzare le eventuali altre funzioni discriminanti oltre le prime due.

1.6. Analisi stepwise

Per effettuare l'analisi discriminante con il metodo stepwise è necessario selezionare "Usa metodo stepwise" nella finestra di dialogo principale. Le opzioni del metodo sono ulteriormente controllabili dalla finestra di dialogo "Metodo stepwise" che si apre con un clic del mouse sul pulsante "Metodo". La tabella 7 presenta i risultati del lambda di Wilks e della F relativa ai passi

dell'analisi. L'ordine di entrata delle variabili nell'equazione per il calcolo delle funzioni discriminanti è il seguente: Coscienziosità (C), Estroversione (E), Nevroticismo (N), Apertura (O), Amicalità (A). La tabella 8 presenta un prospetto delle variabili escluse dall'analisi nei diversi passi in cui si dipana l'analisi. E' facile vedere che ad ogni passo viene inserita la variabile che presenta il valore del coefficiente F più elevato, ovvero quella variabile che contribuisce meglio a differenziare i gruppi. La tabella 9 invece presenta un prospetto relativo alle variabili che ad ogni passo sono inserite nell'equazione: il coefficiente *F di rimozione (F to remove)* viene utilizzato dall'analisi per individuare le variabili che non contribuiscono in modo significativo al differenziare i gruppi. Il coefficiente "*F to remove*" inoltre può essere considerato come ulteriore indice dell'importanza di una variabile indipendente. Un valore elevato di F implica che l'eliminazione della variabile indipendente associata ad esso determinerebbe un decremento sostanziale nella separazione tra i gruppi; alternativamente si può dire che questa variabile influenza in maniera molto forte la variabile dipendente utilizzata per raggruppare i soggetti.

Nessuna variabile viene esclusa, quindi i risultati dell'analisi, quindi la soluzione stepwise finale (in termini di autovalori, devianza spiegata, coefficienti, classificazione, ecc.) è identica alla soluzione ottenuta con il metodo standard.

Tabella 7. Risultati dell'indice lambda di Wilks e della F relativa ai passi dell'analisi.

Variabili inserite/rimosse^{a,b,c,d}

Passo	Inserite	Lambda di Wilks											
		Statistica	df1	df2	df3	F esatto				F approssimato			
						Statistica	df1	df2	Sig.	Statistica	df1	df2	Sig.
1	C	,518	1	3	414,000	128,347	3	414,000	,000				
2	O	,360	2	3	414,000	91,802	6	826,000	,000				
3	N	,259	3	3	414,000					82,794	9	1002,850	,000
4	A	,208	4	3	414,000					73,512	12	1087,695	,000
5	E	,180	5	3	414,000					64,969	15	1132,231	,000

A ogni passo viene inserita la variabile che minimizza il Lambda di Wilks globale.

- a. Il numero massimo di passi è 10
- b. L'F parziale minimo di inserimento è 3.84
- c. L'F parziale massimo di rimozione è 2.71
- d. Livello F, tolleranza o VIN insufficienti. Impossibile eseguire ulteriori calcoli.

Tabella 8. Prospetto delle variabili escluse dall'analisi stepwise nei diversi passi

Variabili non nell'analisi					
Passo		Tolleranza	Tolleranza min.	F di inserimento	Lambda di Wilks
0	N	1,000	1,000	73,135	,654
	E	1,000	1,000	66,397	,675
	O	1,000	1,000	60,636	,695
	A	1,000	1,000	42,555	,764
	C	1,000	1,000	128,347	,518
1	N	,997	,997	49,980	,380
	E	,984	,984	58,404	,364
	O	1,000	1,000	60,509	,360
	A	,991	,991	41,940	,397
2	N	,976	,976	53,764	,259
	E	,975	,975	35,753	,286
	A	,973	,973	33,635	,289
3	E	,949	,949	24,385	,220
	A	,960	,960	33,549	,208
4	E	,945	,935	20,999	,180

Tabella 9. Prospetto delle variabili inserite ad ogni passo nell'equazione

Variabili nell'analisi				
Passo		Tolleranza	F di rimozione	Lambda di Wilks
1	C	1,000	128,347	
2	C	1,000	128,063	,695
	O	1,000	60,509	,518
3	C	,997	98,861	,445
	O	,979	64,486	,380
	N	,976	53,764	,360
4	C	,986	99,980	,359
	O	,965	54,445	,290
	N	,963	53,629	,289
	A	,960	33,549	,259
5	C	,966	100,290	,312
	O	,953	37,054	,229
	N	,935	43,100	,237
	A	,956	29,952	,220
	E	,945	20,999	,208

1.7. Analisi gerarchica

Per effettuare l'analisi gerarchica è necessario controllare l'ordine in cui le variabili sono inserite nell'equazione. Questo non è possibile tramite menù, ma si effettua tramite il linguaggio SINTASSI con il comando "ANALYSIS". In particolare, bisogna assegnare ad ogni variabile indipendente un numero, per cui le variabili con numero più elevato entrano per prime in equazione. Nelle successive linee di comando nel linguaggio SINTASSI l'ordine di inserimento nell'equazione è il seguente: O, E, N, C, A.

*** Controllare l'ordine di inserimento delle variabili.**

```
DISCRIMINANT /GROUPS=clu4_1(1 4)
/VARIABLES= n e o a c
/ANALYSIS n (3) e (4) o (5) a (1) c (2)
/METHOD=WILKS /FIN= 3.84 /FOUT= 2.71
/PRIORS SIZE /HISTORY /PLOT=COMBINED MAP
/CLASSIFY=NONMISSING POOLED.
```

La tabella 10 presenta gli indici di Tolleranza, "F di rimozione" e Lambda di Wilks associati alle variabili nei diversi passi. E' possibile notare che le F associate a tutte le variabili risultano statisticamente significative in ogni passo, dunque ogni variabile contribuisce significativamente nell'equazione discriminante. Considerando quanto emerge nell'ultimo passo dell'analisi gerarchica, ovvero il passo 5, è possibile dire che i risultati finali della soluzione gerarchica sono identici a quelli ottenuti con gli altri due metodi analitici (quello standard discusso nel testo, pagg. 180 e segg., e quello stepwise descritto nel paragrafo precedente).

Tabella 10. Indici statistici della soluzione gerarchica

Variabili nell'analisi				
Passo		Tolleranza	F di rimozione	Lambda di Wilks
1	O	1,000	60,636	
2	O	,991	38,945	,675
	E	,991	44,067	,695
3	O	,966	41,617	,491
	E	,967	24,771	,445
	N	,955	54,439	,526
4	O	,965	41,057	,285
	E	,949	24,385	,259
	N	,950	41,179	,286
	C	,978	98,147	,377
5	O	,953	37,054	,229
	E	,945	20,999	,208
	N	,935	43,100	,237
	C	,966	100,290	,312
	A	,956	29,952	,220

2. MANOVA

2.1. MANOVA con il linguaggio SINTASSI

Di seguito presentiamo i programmi SINTASSI per l'analisi descritta nel testo.

```
**** MANOVA: modello base con test M di Box per
l'omogeneità.
```

```
GLM
  n e o a c BY clu4_1
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /PRINT = HOMOGENEITY
  /CRITERIA = ALPHA(.05)
  /DESIGN = clu4_1 .
```

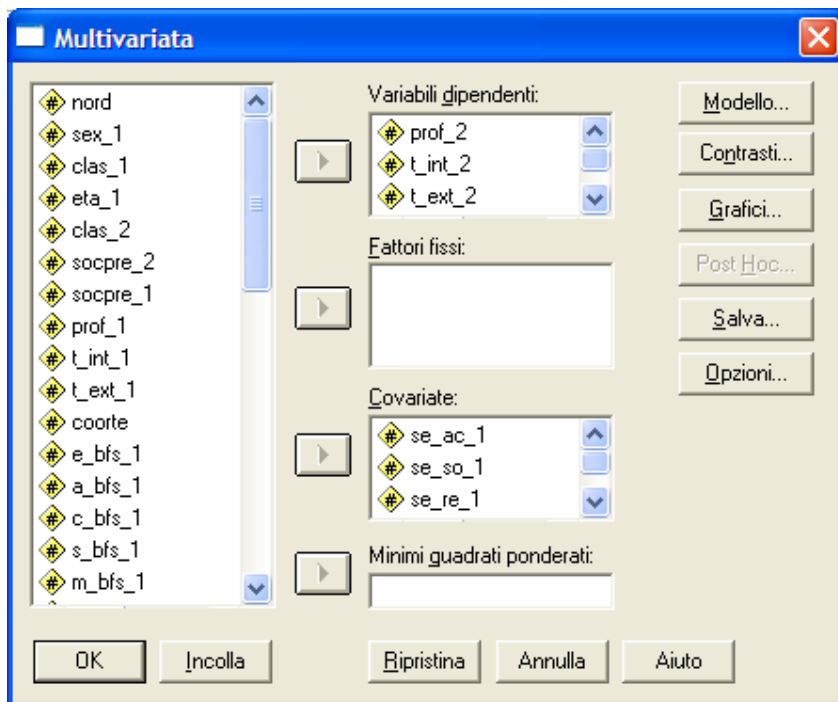

Di fatto ci sono solo due elementi di differenza in questa analisi rispetto a quanto viene riportato nel documento relativo all'analisi della varianza univariata in merito ai disegni entro i soggetti. Un primo elemento riguarda la presenza del comando `/PRINT = HOMOGENEITY` che serve ad ottenere in output il test M di Box per verificare l'omogeneità delle matrici di varianza/covarianza. L'altro elemento di differenza, questo sicuramente più rilevante, è l'assenza del comando `"/WSFACTOR"` utilizzato nell'ANOVA per definire il fattore o i fattori entro i soggetti. In assenza di tale comando il programma interpreta il disegno come una analisi della varianza multivariata e non come un disegno con un fattore entro i soggetti (o "a prove ripetute").

2.2. L'Analisi Multivariata della Covarianza (MANCOVA)

L'approccio della MANOVA può essere combinato con l'approccio dell'Analisi della Covarianza, in quella che viene definita Analisi Multivariata della Covarianza (MANCOVA). Si tratta di un disegno utilizzato non molto spesso ma che presenta delle interessanti potenzialità. Da un lato può essere utilizzato per rimuovere, in una MANOVA, l'effetto indesiderato di alcune variabili di disturbo che vengono inserite appunto come covariate. Dall'altro può essere utilizzato per effettuare una *regressione multipla multivariata*, ovvero un modello di regressione multipla in cui si esamina simultaneamente l'impatto di un insieme di variabili indipendenti su un insieme di variabili dipendenti (la regressione multipla infatti considera l'impatto di un insieme di variabili indipendenti su *una sola* variabile dipendente).

Nella figura 3 presentiamo la finestra di dialogo iniziale della procedura GLM che consente di esaminare in SPSS un modello di regressione multipla multivariata. E' possibile notare che nessuna variabile viene inserita nella lista dei fattori fissi, e quindi il modello analizza l'impatto che le covariate hanno sulle variabili dipendenti.

Figura 3. Finestra di dialogo per l'analisi multivariata della covarianza



Nell'esempio di questo tipo di applicazione della MANCOVA abbiamo considerato l'impatto simultaneo delle 3 variabili relative all'autoefficacia percepita utilizzate in un esempio del secondo capitolo (autoefficacia scolastica, sociale e regolativa) sulle tre variabili indipendenti descritte sempre nel secondo capitolo (internalizzazione, esternalizzazione, profitto scolastico). Si tratta di un esempio di analisi descritto nel testo di Barbaranelli (2003) nel paragrafo 5.2.6.

La tabella 11 presenta l'output relativo agli effetti multivariati. Tutte le tre variabili indipendenti presentano un impatto significativo sulle tre variabili dipendenti prese nel loro complesso, come evidenziano i valori del coefficiente Lambda di Wilks. Come si ricava dall'indice eta quadrato, la variabile più importante è l'efficacia scolastica, seguita dall'efficacia regolativa ed infine dall'efficacia sociale. La tabella 12 presenta gli effetti univariati fra i soggetti. Diversamente dagli effetti multivariati, si tratta di effetti univariati che esaminano la percentuale di varianza che di ogni

variabile dipendente presa separatamente viene spiegata dalle 3 variabili indipendenti, e la significatività dell'impatto di ciascuna delle indipendenti su ciascuna delle dipendenti. La parte finale della tabella riporta la proporzione di varianza di ciascuna singola variabile dipendente che è spiegata dalla regressione.

La stima dell'impatto delle variabili indipendenti sulle singole variabili dipendenti è invece evidenziata nella tabella 13 che appunto è dedicata alla stima univariata dei parametri. Mentre l'impatto della efficacia scolastica e della efficacia regolativa risulta sempre significativo anche nel test univariato (con una prevalenza dell'efficacia scolastica sull'efficacia regolativa), l'impatto dell'efficacia sociale non è mai significativo nei tre modelli univariati: questo conferma quanto evidenziato nella tabella 12, mentre è in disaccordo con quanto evidenziato nel test multivariato in tabella 11. A tale riguardo è possibile concludere che la MANCOVA risulta più potente rispetto alle singole regressioni multiple univariate, nell'evidenziare un'influenza significativa (anche se debole) dell'efficacia sociale sulle tre variabili dipendenti esaminate complessivamente. Questa maggiore potenza emerge chiaramente se si considerano i valori della potenza osservata nelle tabelle 11 e 12.

Tabella 11. Effetti multivariati

Test multivariat ^f									
Effetto		Valore	F	Ipotesi df	Gradi di libertà dell'errore	Sig.	Eta quadrato parziale	Non centralità	Potenza osservata ^a
Intercetta	Traccia di Pillai	,186	34,274 ^b	3,000	449,000	,000	,186	102,822	1,000
	Lambda di Wilks	,814	34,274 ^b	3,000	449,000	,000	,186	102,822	1,000
	Traccia di Hotelling	,229	34,274 ^b	3,000	449,000	,000	,186	102,822	1,000
	Radice di Roy	,229	34,274 ^b	3,000	449,000	,000	,186	102,822	1,000
SE_AC_1	Traccia di Pillai	,108	18,157 ^b	3,000	449,000	,000	,108	54,471	1,000
	Lambda di Wilks	,892	18,157 ^b	3,000	449,000	,000	,108	54,471	1,000
	Traccia di Hotelling	,121	18,157 ^b	3,000	449,000	,000	,108	54,471	1,000
	Radice di Roy	,121	18,157 ^b	3,000	449,000	,000	,108	54,471	1,000
SE_SO_1	Traccia di Pillai	,021	3,159 ^b	3,000	449,000	,025	,021	9,476	,732
	Lambda di Wilks	,979	3,159 ^b	3,000	449,000	,025	,021	9,476	,732
	Traccia di Hotelling	,021	3,159 ^b	3,000	449,000	,025	,021	9,476	,732
	Radice di Roy	,021	3,159 ^b	3,000	449,000	,025	,021	9,476	,732
SE_RE_1	Traccia di Pillai	,069	11,138 ^b	3,000	449,000	,000	,069	33,414	,999
	Lambda di Wilks	,931	11,138 ^b	3,000	449,000	,000	,069	33,414	,999
	Traccia di Hotelling	,074	11,138 ^b	3,000	449,000	,000	,069	33,414	,999
	Radice di Roy	,074	11,138 ^b	3,000	449,000	,000	,069	33,414	,999

a. Calcolato usando alfa = ,05

b. Statistica esatta

c. Disegno: Intercept+SE_AC_1+SE_SO_1+SE_RE_1

Tabella 12. Effetti univariati fra i soggetti

Test degli effetti fra soggetti

Sorgente	Variabile dipendente	Somma dei quadrati Tipo III	df	Media dei quadrati	F	Sig.	Eta quadrato parziale	Non centralità	Potenza osservata ^a
Modello corretto	PROF_2	79,344 ^b	3	26,448	42,818	,000	,222	128,453	1,000
	T_INT_2	1,019 ^c	3	,340	14,164	,000	,086	42,493	1,000
	T_EXT_2	3,600 ^d	3	1,200	22,059	,000	,128	66,177	1,000
Intercetta	PROF_2	,122	1	,122	,198	,657	,000	,198	,073
	T_INT_2	1,577	1	1,577	65,762	,000	,127	65,762	1,000
	T_EXT_2	3,113	1	3,113	57,211	,000	,113	57,211	1,000
SE_AC_1	PROF_2	33,700	1	33,700	54,559	,000	,108	54,559	1,000
	T_INT_2	,129	1	,129	5,378	,021	,012	5,378	,638
	T_EXT_2	,841	1	,841	15,462	,000	,033	15,462	,975
SE_SO_1	PROF_2	1,309	1	1,309	2,119	,146	,005	2,119	,306
	T_INT_2	5,477E-02	1	5,477E-02	2,284	,131	,005	2,284	,326
	T_EXT_2	7,264E-02	1	7,264E-02	1,335	,249	,003	1,335	,211
SE_RE_1	PROF_2	14,252	1	14,252	23,073	,000	,049	23,073	,998
	T_INT_2	,248	1	,248	10,343	,001	,022	10,343	,894
	T_EXT_2	1,452	1	1,452	26,697	,000	,056	26,697	,999
Errore	PROF_2	278,578	451	,618					
	T_INT_2	10,816	451	2,398E-02					
	T_EXT_2	24,537	451	5,441E-02					
Totale	PROF_2	3872,496	455						
	T_INT_2	21,661	455						
	T_EXT_2	46,856	455						
Totale corretto	PROF_2	357,923	454						
	T_INT_2	11,835	454						
	T_EXT_2	28,138	454						

a. Calcolato usando alfa = ,05

b. R quadrato = ,222 (R quadrato corretto = ,217)

c. R quadrato = ,086 (R quadrato corretto = ,080)

d. R quadrato = ,128 (R quadrato corretto = ,122)

Tabella 13. Stime univariate (non standardizzate) dei parametri

Stime dei parametri

Variabile dipendente	Parametro	B	Errore std.	t	Sig.	Intervallo di confidenza 95%	
						Limite inferiore	Limite superiore
PROF_2	Intercetta	,149	,335	,445	,657	-,509	,806
	SE_AC_1	,028	,004	7,386	,000	,021	,036
	SE_SO_1	-,010	,007	-1,456	,146	-,024	,004
	SE_RE_1	,052	,011	4,803	,000	,031	,073
T_INT_2	Intercetta	,535	,066	8,109	,000	,405	,664
	SE_AC_1	-,002	,001	-2,319	,021	-,003	,000
	SE_SO_1	-,002	,001	-1,511	,131	-,005	,001
	SE_RE_1	-,007	,002	-3,216	,001	-,011	-,003
T_EXT_2	Intercetta	,751	,099	7,564	,000	,556	,946
	SE_AC_1	-,004	,001	-3,932	,000	-,007	-,002
	SE_SO_1	,002	,002	1,155	,249	-,002	,006
	SE_RE_1	-,017	,003	-5,167	,000	-,023	-,010

a. Calcolato usando alfa = ,05

Per ottenere i coefficienti standardizzati è necessario moltiplicare il coefficiente per il rapporto delle deviazioni standard della variabile indipendente e della dipendente, ovvero $\beta_{y,x}=b_{y,x} s_x/s_y$. La tabella 14 presenta il risultato di questa operazione effettuata tramite il programma Excel.

Tabella 14. Stime univariate (standardizzate) dei parametri

Variabile dipendente	Variabile indipendente	B	dsy	dsx	beta
PROF_2	SE_AC_1	0,028	0,888	12,108	0,384
	SE_SO_1	-0,010	0,888	6,549	-0,074
	SE_RE_1	0,052	0,888	3,742	0,218
T_INT_2	SE_AC_1	-0,002	0,161	12,108	-0,131
	SE_SO_1	-0,002	0,161	6,549	-0,084
	SE_RE_1	-0,007	0,161	3,742	-0,158
T_EXT_2	SE_AC_1	-0,004	0,249	12,108	-0,217
	SE_SO_1	0,002	0,249	6,549	0,062
	SE_RE_1	-0,017	0,249	3,742	-0,249

Infine presentiamo di seguito il programma SINTASSI per la MANCOVA

```
**** MANCOVA.
```

```
GLM
```

```
prof_2 t_int_2 t_ext_2 WITH se_ac_1 se_so_1 se_re_1
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/PRINT = PARAMETER ETASQ OPOWER
/CRITERIA = ALPHA(.05)
/DESIGN = se_ac_1 se_so_1 se_re_1 .
```

E' da notare che nel comando GLM non viene specificata la parola "BY" quindi non ci sono variabili tra i soggetti, ma solo le 3 covariate che seguono "WITH". Questo è ulteriormente specificato nel comando "/DESIGN" dove gli effetti che vengono esaminati sono solo quelli delle covariate. Inoltre, poiché manca il comando "/WSFACTOR" le tre variabili dipendenti che seguono il comando GLM vengono interpretate come variabili dipendenti "separate" e non come livelli di un

fattore entro i soggetti. Infine, le opzioni ETASQ OPOWER di “/PRINT” vengono specificate per ottenere nell’output rispettivamente le stime dell’ampiezza dell’effetto e della potenza osservata.

Riferimenti bibliografici

Barbaranelli, C. (2003). *Analisi dei dati*. Milano: LED