

## Analisi delle Corrispondenze Multiple Prof. Roberto Fantaccione

Consideriamo il nostro dataset formato da 468 individui e 12 variabili nominali costituite dalle seguenti modalità :

colonna D: Age of client	(età del cliente)	-> 4 modalità (meno di 23;da 23 a 40;da 40 a 50;oltre 50)
colonna E: Profession	(professione)	-> 3 modalità (dipendente ;esecutivo; altra professione)
colonna F: Seniority	(anzianità)	-> 5 modalità( 1 anno o meno;da 1 a 4 anni;da 4 a 6;da 6 a 12;più di 12)
colonna G: Family Situation	(situazione di famiglia)	-> 4 modalità ( single;sposato;divorziato;vedovo)
colonna H: Type of client	(tipo di cliente)	-> 2 modalità ( buono; cattivo)
colonna I: Home of employee	(a casa del dipendente)	-> 2 modalità (stipendio a domicilio e non)
colonna J: Size of savings	(dimensione dei risparmi)	-> 4 modalità(nessun risparmio;meno di 10 KF;da 10 a 100 kf; oltre 100 KF)
colonna K: Active mean migliaia di franchi	(media attiva in ml di franchi)	->3 modalità (meno di 2KF;da 2 a 5;più di 5)
colonna L: Mean of movements	(media del movim. di trasl.)	-> 4 modalità (meno di 10KF;da 10 a 30;da 30 a 50;più di 50)
colonna M: Cumulative debits	(debiti cumulativi)	-> 3 modalità (meno di 40;da 40 a 100;più di 100)
colonna N: overdraft authorized	(scoperto autorizzato)	-> 2 modalità (proibito;autorizzato)
colonna O: Checkbook not allowed	(assegni ammessi)	-> 1 modalità (libretto di assegni ammessi)

Totale 37 modalità

**Figura 1 – Matrice di partenza**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	KIDEN	Osservazione	Classe	Age of client	Profession	Seniority	Family Situation	Type of client	Home of employee	Size of savings	Active mean migliaia	Mean of movement	Cumulative debits	overdraft authorized	Chec
2	1	1	1	over 50 years	employee	plus 12 years	single	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	from 40 to 100	forbidden	chec
3	2	2	1	less than 23 years	employee	1 year or less	single	good client	domicile salary	no savings	from 2 to 5 KF	Less than 10 KF	less than 40	forbidden	chec
4	3	3	1	from 23 to 40 years	employee	from 6 to 12 years	widow	bad client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	from 40 to 100	authorised	chec
5	4	5	1	less than 23 years	employee	from 6 to 12 years	single	good client	not domicile salary	no savings	from 2 to 5 KF	From 10 to 30 KF	less than 40	forbidden	chec
6	5	6	1	from 23 to 40 years	employee	1 year or less	single	good client	domicile salary	no savings	from 2 to 5 KF	Less than 10 KF	less than 40	authorised	chec
7	6	7	1	over 50 years	executive	from 6 to 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	less than 40	forbidden	chec
8	7	8	1	over 50 years	executive	plus 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	less than 40	forbidden	chec
9	8	10	1	over 50 years	employee	from 4 to 6 years	single	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	from 40 to 100	forbidden	chec
10	9	11	1	over 50 years	employee	plus 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	from 40 to 100	forbidden	chec
11	10	12	1	from 40 to 50 years	executive	1 year or less	married	good client	not domicile salary	less than 10 KF	from 2 to 5 KF	From 30 to 50 KF	less than 40	forbidden	chec
12	11	13	1	from 23 to 40 years	other profession	from 4 to 6 years	single	good client	not domicile salary	no savings	from 2 to 5 KF	Less than 10 KF	from 40 to 100	authorised	chec
13	12	14	1	from 23 to 40 years	employee	from 6 to 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	from 40 to 100	authorised	chec
14	13	16	1	from 40 to 50 years	employee	from 6 to 12 years	divorced	bad client	domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	from 40 to 100	authorised	chec
15	14	17	1	over 50 years	other profession	plus 12 years	single	good client	domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	less than 40	forbidden	chec
16	15	18	1	over 50 years	other profession	plus 12 years	widow	bad client	domicile salary	no savings	from 2 to 5 KF	Less than 10 KF	from 40 to 100	authorised	chec
17	16	19	1	over 50 years	employee	from 4 to 6 years	married	good client	domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	less than 40	authorised	chec
18	17	21	1	from 40 to 50 years	employee	from 1 to 4 years	married	good client	domicile salary	less than 10 KF	from 2 to 5 KF	Less than 10 KF	from 40 to 100	authorised	chec
19	18	26	1	from 40 to 50 years	employee	1 year or less	married	good client	not domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	less than 40	authorised	chec
20	19	28	1	from 23 to 40 years	employee	from 4 to 6 years	married	good client	domicile salary	no savings	from 2 to 5 KF	From 10 to 30 KF	from 40 to 100	authorised	chec
21	20	30	1	over 50 years	employee	from 6 to 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	from 40 to 100	forbidden	chec
22	21	31	1	from 40 to 50 years	other profession	1 year or less	single	good client	domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	from 40 to 100	forbidden	chec
23	22	33	1	from 40 to 50 years	employee	plus 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	from 40 to 100	authorised	chec
24	23	37	1	from 40 to 50 years	executive	plus 12 years	divorced	good client	domicile salary	no savings	from 2 to 5 KF	more than 50 KF	less than 40	forbidden	chec
25	24	39	1	from 23 to 40 years	employee	1 year or less	single	good client	domicile salary	no savings	from 2 to 5 KF	Less than 10 KF	from 40 to 100	authorised	chec
26	25	41	1	from 40 to 50 years	employee	plus 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	less than 40	forbidden	chec
27	26	43	1	over 50 years	executive	from 6 to 12 years	married	good client	domicile salary	no savings	from 2 to 5 KF	From 30 to 50 KF	from 40 to 100	forbidden	chec

Eseguiamo una statistica semplice per  
Il calcolo delle frequenze relative delle rispettive modalità

Statistiche semplici:			
Variabile	Modalità	frequenz	%
Age of client	from 23 to 40 years	150	32,051
	from 40 to 50 years	122	26,068
	less than 23 years	88	18,803
	over 50 years	108	23,077
Profession	employee	237	50,641
	executive	77	16,453
	other profession	154	32,906
Seniority	1 year or less	193	42,521
	from 1 to 4 years	47	10,043
	from 4 to 6 years	69	14,744
	from 6 to 12 years	66	14,103
Family Situation	plus 12 years	87	18,590
	divorced	61	13,034
	married	221	47,222
	single	170	36,325
Type of client	widow	16	3,419
	bad client	231	49,359
	good client	237	50,641
	Home of employee	domicile salary	316
Size of savings	not domicile salary	152	32,479
	from 10 to 100 KF	32	6,838
	less than 10 KF	58	12,393
	more than 100 KF	8	1,709
Active mean migliaia di franchi	no savings	370	79,060
	from 2 to 5 KF	308	65,812
	less than 2 KF	98	20,940
	more than 5 KF	62	13,248
Mean of movements	From 10 to 30 KF	71	15,171
	From 30 to 50 KF	129	27,564
	Less than 10 KF	154	32,906
	more than 50 KF	114	24,359
Cumulative debits	from 40 to 100	161	34,402
	less than 40	171	36,538
	more than 100	136	29,060
overdraft authorized	authorised	202	43,162
	forbidden	266	56,838
Checkbook not allowed	checkbook allowed	468	100,000





Lavorando con variabili categoriali non si può ragionare usando la metrica di tipo Euclideo, ma si dovrà usare la metrica del Chi-quadrato. Le distanze si calcolano tra i profili riga (o colonna).

Il calcolo della distanza del chi-quadrato inserisce un coefficiente di ponderazione che rivaluta le componenti a più bassa frequenza e ridimensiona quelle a più alta frequenza. Il coefficiente di ponderazione T/Cj è l'inverso del rapporto Cj/T che è la frequenza relativa marginale della modalità J: si pondera ogni elemento per l'inverso della sua importanza sul totale delle frequenze)

$$D^2 (ir) = \sum_{j=1}^q T/C_j (f_{ij}/F_i - fr_j/F_r)^2$$

### Obiettivo ACM

Trovare poche dimensioni in grado di riprodurre la maggior parte dell'associazione presente fra le variabili analizzate in un ridotto numero di fattori. E' possibile scegliere quali variabili far partecipare alla determinazione dei fattori (variabili attive). Le variabili che non partecipano all'elaborazione possono essere analizzate in qualità di variabili illustrative (controllando la loro relazione con i fattori ottenuti con l'ACM)

I fattori ottenuti con l'ACM:

- sono ortogonali fra loro
- sono combinazioni delle variabili-modalità inserite nell'analisi
- riproducono, in ordine decrescente, il massimo della dispersione (o inerzia) presente nella matrice di Burt

### Inerzia totale e tassi di inerzia

La matrice Z, di dimensioni (n,s), identifica, nello spazio R<sup>s</sup>, s punti le cui coordinate sono le righe della matrice D<sup>-1</sup>Z' (o le colonne della matrice ZD<sup>-1</sup>). Ciascuna sottomatrice Z<sub>q</sub> genera, in R<sup>s</sup>, un sottospazio lineare a p<sub>q</sub> dimensioni che avrà certamente in comune con tutti gli altri almeno la prima bisettrice, ossia il vettore le cui componenti risultano tutte uguali a uno. Il rango massimo di ZD<sup>-1</sup>, e quindi di Z, è quindi pari a:

$$s_1 + (s_2 - 1) + (s_3 - 1) + \dots + (s_p - 1) = s - p + 1$$

D'altra parte, degli s-p+1 autovettori della matrice Z'ZD<sup>-1</sup>, uno sarà quello che congiunge l'origine con il baricentro e che corrisponde all'autovalore banale unitario. Nell'analisi centrata, quindi, potranno essere determinati al massimo s-p auto valori non nulli e la percentuale di variabilità spiegata da ciascun fattore sarà rappresentata dalla quantità:

Tassi di inerzia :

$$\frac{\lambda_{\alpha}}{\sum_{j=1}^{s-p} \lambda_{\alpha}} \times 100$$

Tale quantità, che è una misura eccessivamente pessimistica dell'effettivo potere esplicativo dei diversi fattori, assume, nel caso dell'ACM, valori generalmente molto bassi anche per i primi fattori. Questo perché "la codifica disgiuntiva, imponendo una relazione di ortogonalità tra le modalità di una stessa variabile, introduce una sorta di sfericità artificiale della nube dei punti", che si manifesta proprio nei bassi valori dei tassi di inerzia.

L'inerzia di una modalità, l(j), è data dal quadrato della distanza dal baricentro per il peso m<sub>j</sub>, rappresentato dal corrispondente marginale di colonna sulla tabella delle frequenze relative:

$$m_j = \frac{z_{\cdot j}}{n \cdot p}$$

Nel nostro caso

Inerzia totale: 2,083

Gli s-p=25 autovalori non nulli che è possibile determinare sono i seguenti:

### Autovalori e tassi di inerzia

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24	F25	
Autovalore	0,249	0,168	0,143	0,131	0,108	0,104	0,100	0,091	0,089	0,084	0,080	0,078	0,074	0,072	0,067	0,064	0,061	0,058	0,051	0,051	0,046	0,041	0,034	0,027	0,011	
Inerzia (%)	11,933	8,055	6,953	6,290	5,207	5,004	4,795	4,348	4,293	4,020	3,850	3,725	3,567	3,457	3,215	3,069	2,948	2,795	2,467	2,439	2,223	1,968	1,656	1,305	0,519	
% cumulata	11,933	19,988	26,841	33,131	38,337	43,341	48,137	52,485	56,778	60,798	64,648	68,373	71,940	75,397	78,612	81,680	84,628	87,423	89,890	92,329	94,552	96,520	98,176	99,481	100,000	
Inerzia corretta	0,033	0,008	0,004	0,003	0,001	0,001	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	
Inerzia corretta (%)	49,881	13,032	6,452	4,156	1,154	0,799	0,502	0,096	0,068	0,000																
% cumulata	49,881	62,914	69,366	73,522	74,676	75,474	75,976	76,072	76,140																	

Tra i metodi per correggere i valori dei tassi di inerzia e renderli attendibili, soprattutto relativamente ai primi fattori, Benzècri ha introdotto la correzione:

$$\lambda^* = \left(\frac{p}{p-1}\right)^2 \times \left(\lambda - \frac{1}{p}\right)^2 \quad \text{per } \lambda > 1/p$$

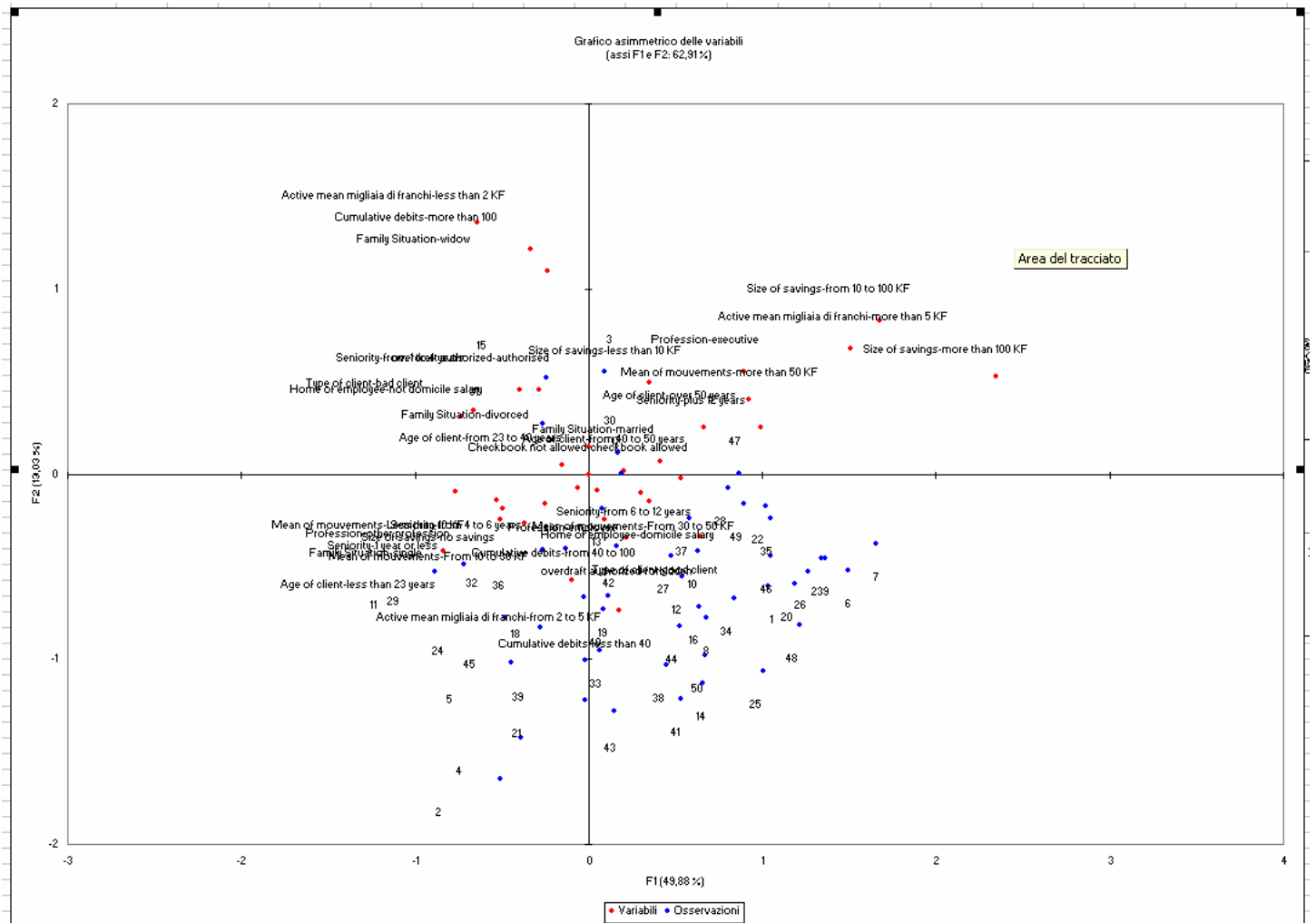
Che, per p elevato, equivale a considerare il quadrato degli autovalori mentre per p=2 porta all'espressione  $\lambda^* = (2\lambda - 1)^2 = \mu$  che individua gli auto valori dell'AC sulla tabella di contingenza che incrocia le due variabili.





Valori-test (Variabili)	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Age of client-from 23 to 40 years	-2,282	0,795	-7,942	-5,390	0,450	-11,928	1,166	-2,826	2,433	-0,121
Age of client-from 40 to 50 years	2,644	0,191	2,430	3,000	-14,587	3,220	3,290	-1,734	-5,066	4,089
Age of client-less than 23 years	-8,680	-4,316	-5,418	5,393	7,341	7,832	0,972	6,769	1,586	-2,046
Age of client-over 50 years	7,823	2,967	11,290	-2,157	7,892	2,593	-5,620	-1,341	1,112	-2,229
Profession-employee	1,109	-1,806	-6,322	-9,995	-2,197	-3,671	0,296	2,626	-1,713	3,234
Profession-executive	8,595	5,317	7,253	3,910	6,821	-1,633	-3,325	3,165	-1,969	1,732
Profession-other profession	-7,962	-2,168	1,004	7,549	-3,045	5,194	2,308	-5,291	3,376	-4,808
Seniority-1 year or less	-9,244	-3,420	2,412	10,451	-3,287	-3,948	-0,686	-2,546	2,009	6,013
Seniority-from 1 to 4 years	-2,857	3,306	-2,156	-2,836	-1,562	1,590	-1,821	-6,739	4,796	-13,268
Seniority-from 4 to 6 years	-0,551	-0,705	-3,451	-5,421	0,195	4,172	-1,603	13,617	-5,738	0,896
Seniority-from 6 to 12 years	4,677	-0,223	-7,712	-3,543	7,203	-4,250	0,960	-6,120	-5,875	3,159
Seniority-plus 12 years	10,272	2,634	8,646	-2,981	-1,238	3,789	2,862	1,598	4,227	-1,024
Family Situation-divorced	-0,014	1,221	1,862	1,883	-3,462	-7,724	-14,817	-0,430	-3,871	3,393
Family Situation-married	8,378	1,365	5,314	-1,392	-7,674	-1,475	8,603	-1,948	-2,041	-5,247
Family Situation-single	-8,328	-3,961	-9,824	0,403	7,302	4,446	2,884	3,840	4,436	2,471
Family Situation-widow	-0,948	4,470	7,952	-0,732	8,171	6,596	-3,813	-4,017	1,039	1,586
Type of client-bad client	-14,142	7,378	1,998	3,136	-0,579	-1,372	-1,161	-3,576	-2,254	1,678
Type of client-good client	14,142	-7,378	-1,998	-3,136	0,579	1,372	1,161	3,576	2,254	-1,678
Home of employee-domicile salary	10,962	-4,640	-1,467	-6,082	-1,252	-2,780	1,467	1,698	6,096	-0,132
Home of employee-not domicile salary	-10,962	4,640	1,467	6,082	1,252	2,780	-1,467	-1,698	-6,096	0,132
Size of savings-less than 10 KF	9,794	4,867	-10,311	6,066	1,266	3,570	-2,225	-2,473	-8,051	-5,819
Size of savings-more than 10 KF	2,864	4,006	-4,277	-0,134	-3,709	1,479	-9,329	1,017	13,291	5,689
Size of savings-no savings	6,693	1,506	-0,339	8,532	-0,095	-1,915	7,851	2,087	5,822	4,387
Active mean migliaia di franchi-less than 2 K€	-2,990	-17,081	5,857	-3,873	0,656	0,343	-1,085	-3,027	-2,372	1,075
Active mean migliaia di franchi-more than 2 K€	-7,132	15,124	2,193	-3,119	-0,114	-2,969	4,108	4,425	1,873	-0,460
Active mean migliaia di franchi-less than 5 K€	12,744	5,748	-10,828	9,162	-0,779	2,244	-3,412	-1,075	1,071	-0,852
Active mean migliaia di franchi-more than 5 K€	-3,380	-2,422	-3,119	-3,571	7,227	-5,988	2,330	-5,476	3,912	0,887
Mean of movements-From 10 to 30 K€	4,050	-1,364	-1,401	-9,482	-6,803	6,696	-4,718	1,345	0,546	-2,475
Mean of movements-From 30 to 50 K€	-11,619	-1,395	1,311	7,812	-2,958	0,307	-2,941	3,129	-0,557	-1,064
Mean of movements-Less than 10 KF	11,328	4,972	2,630	4,304	4,280	-2,368	6,074	-0,249	-3,228	3,668
Mean of movements-more than 10 KF	1,431	-3,866	-2,319	-3,740	-0,510	10,785	1,549	-9,295	-0,978	9,728
Cumulative debits-less than 40	2,904	-12,058	2,879	5,066	1,397	-7,643	-4,218	5,232	-0,241	-7,184
Cumulative debits-more than 40	-4,576	16,834	-0,627	-1,480	-0,949	-3,178	2,853	4,176	1,279	-2,559
overdraft authorized-authorized	-5,388	8,529	-3,474	-7,610	-0,523	3,124	-0,770	-1,047	-2,665	2,721
overdraft authorized-forbidden	5,388	-8,529	3,474	7,610	0,523	-3,124	0,770	1,047	2,665	-2,721
Checkbook not allowed-checkbook allowed										

Figura 6 – Piano fattoriale 1-2: modalità attive (prodotto con software XLSTAT)



La contrapposizione delle modalità sugli assi è ben visibile, in alto abbiamo i clienti che hanno un'età superiore a 50 anni che presentano le seguenti caratteristiche:

- hanno una media di movimenti superiore ai 50.000 franchi, con un'anzianità superiore a 12 anni e con stato familiare coniugato. Possiamo affermare che queste sono le categorie che risparmiano di più ed hanno una media attiva di oltre 5000 franchi, inoltre, la loro professione è maggiormente executive (dirigenti).
- Nella parte centrale del piano fattoriale di fig. 6, abbiamo le categorie tra i 23 e 40 anni, esse risultano essere maggiormente divorziati e per la banca non sono buoni clienti.
- I clienti con età compresa tra 40 e 50 si trovano in una situazione intermedia rispetto ai due gruppi analizzati.
- Nella parte bassa a sinistra abbiamo i clienti con età al di sotto di 23 anni. Essi risultano essere per la maggior parte Single ed hanno una media attiva in franchi compresa tra 2 e 5 KF, anche perché sono clienti da meno di 12 mesi con cumulo di debiti meno di 40 mila franchi.

Analizziamo ora il piano formato da terzo e quarto fattore:

Si evince che questo piano contribuisce a definire il profilo dei soggetti analizzati e mette in risalto aspetti quali le professioni e più precisamente:

- i clienti sotto i 23 anni non sono dipendenti della banca. Da una prima analisi troviamo pochi individui appartenenti a questa categoria.

Un aspetto importante è che pochi individui risparmiano più di 100.000 franchi.

Il tutto conferma che il correntista più anziano ha una buona condizione economica e risulta essere un buon cliente. Tale fattore spiega il fatto che la banca tende a mantenere i rapporti con clienti buoni, in tal modo abbassa il rischio di inadempienza di un debitore (cliente cattivo) e tiene sotto controllo la riserva o il proprio patrimonio di Vigilanza così come stabilito da Basilea 2.

Analisi eseguita dal Prof. Roberto Fantaccione



Grafico asimmetrico delle variabili  
(assi F3 e F4: 10,61 %)

